

A research council for Europe?

Everybody is anxious to help mould Europe's research policy. The place to start is to say that episodic funding (by means of 'framework' programmes) is a bad idea. What is needed is a truly European research council.

THE Royal Society of London has done a public service by making public its response to a request from the European Science Foundation (ESF) for opinions on the European Commission's Fifth Framework Programme, now being worked up in Brussels to cover the period 1997–2001. ESF's slightly subversive plan is that it should collect the views of its member organizations, national research councils and academies, and deliver an amalgamated opinion to the commission in the spring of next year. Unintentionally but helpfully, the Royal Society's opinion reflects national ambivalence about the now-substantial volume of research supported from Brussels, much of which is committed to projects in applied science, as in telecommunications. Logic would suggest that Brussels should be more active in research, but suspicion of the commission's procedures and hankering after the doctrine of subsidiarity (that the centre should be responsible only for what national organizations cannot do) makes the Royal Society pull back from that conclusion.

The chief of the society's worries is that basic research will be short-changed, for which reason it asks that 10 per cent of the budgets of individual projects should be set aside for basic research, but spent through a different mechanism. Referring to the present procedures for project selection and the award of grants, the society pleads for greater transparency. It argues that there are some fields in which the commission could usefully subcontract to other organizations the business of recommending the award of research grants. It asks that there should be regular mid-term reviews of the commission's major projects, and that the European Science and Technology Assembly should play a central part in responding to these evaluations. If followed, these recommendations would make for a better-managed Fifth Framework Programme. But they fall short of being what the commission needs to be told.

Whether or not Europe becomes a formal federation of its member states, it is only a matter of time before there has to be a European research council of some kind. At the least, there will be a need to make grants for infrastructure development or for making sure that young people from countries whose habit is to spend too little on science are nevertheless caught up in the swim. (The Royal Society has sensible proposals on both counts.) So why not make a start by creating such a grant-making agency here and now, with the intention that when (not *if*) it has proved its worth, its terms of reference will be

extended. The recipe for such an agency is easily defined: a supervisory board that commands respect among working scientists (but whose composition will be rotated regularly), an impartial secretariat recruited by merit (usually experience) and a budget transferred from the commission in which the donors will have no further say.

While the functioning of such an agency is easily described, it is a fair bet that its routine operations would not instantly work well. Two or three years might go by before it settled down. Then it would be possible to think of extending its terms of reference, which would of course require the compliance of the Council of Ministers (on which national governments are represented), no doubt influenced and informed by the constitutional direction Europe is then taking. Inevitably, there would be petty arguments (which would have to be dismissed) about questions such as whether the ultimate contributors to the funds were deriving benefits in proportion to the sums handed over.

At this early stage (in 1995), it is too soon even to guess at what might become of such an agency. It might never be more than a mechanism for infrastructure grants and for helping countries not doing enough to help themselves. But if, in an ideal world, such an agency became a means of making grants to good people without their having to promise that they will engage Europeans of other nationalities in their projects, Europe could find that it has a powerful motor of European research at its service. As things are, national research councils hug to themselves the most imaginative proposals steered in their direction. It will be a state of grace when they can put the shoe on the other foot, seeing whether they can hive off the most imaginative proposals to the central agency, reckoning that, if they succeed, they will have more to spend on the national interest.

Naturally it will not be easy to get from here to there. But that is all the more reason for beginning soon. The Royal Society seems, from its judicious comment on the Fifth Framework Programme, to be inclined in that direction. It will be no surprise if other research councils and academies are similarly disposed. In that case, ESF should give the venture a hearing. And certainly it should say in its advice next spring that the time has come to abandon the 'framework' concept. Other functions of the European enterprise are funded annually. Why should research be different? □



Global warming rows

The Intergovernmental Panel on Climate Change must reorganize its work-programme in everybody's interest.

THE global warming business is in danger of getting out of hand, for which reason it is necessary to restate the position on this issue that *Nature* has taken for several years: the accumulation of carbon dioxide and other greenhouse gases in the Earth's atmosphere means that there is a high chance, perhaps 80 per cent or so, that there will be an increase of average surface temperature of about one degree or thereabouts in the next half century. Crucially, the statement is not equivalent to the statement that there will be an increase of average temperature equal to roughly 80 per cent of one degree, or of 0.8 degrees. Rather, it is that it is more probable (by a factor of four or thereabouts) that there will be an increase of temperature than that there will be none, but the latter cannot be entirely ruled out.

Despite the qualifications, the prospect is a great threat to our present way of life: if the global warming scenario is proved to be correct, patterns of agricultural productivity will be greatly changed, some now-prosperous countries will be endangered or even extinguished by rising sea levels and everybody will know that the hazards thus visited upon them will be merely a beginning. If, 50 years from now, the average temperature has increased by one degree, it will keep on rising afterwards unless something has been done to abate the greenhouse effect substantially. So even in its qualified form, *Nature's* position on global warming, which is widely echoed in the scientific community, demands great things of politicians. The 1990 Rio treaty was a step in the right direction, but will not suffice if global warming proves to be a reality.

All that is by way of apologia to, or of pre-emptive defence against, the international body called the Intergovernmental Panel on Climate Change (IPCC), which calls the shots on global warming and which is hypersensitive to any suggestion that its doings fall short of perfection in any way. Two weeks ago (*Nature* 378, 119; 1995), it emerged that one of IPCC's three working groups had decided to publish a summary of a chapter in a forthcoming technical report with which the authors disagreed so sharply that they were planning to withhold copyright consent. This week (see page 329), a committee of the US Congress has been told that bona fide scientists had been denied access to data in the possession of IPCC (and subsequently published in *Nature*).

In the calendar of human calumny, neither incident looms large. The trouble about the chapter whose meaning was distorted by its summary turns on an attempt to put a value on an average human life (in US dollars), which would then relate to the compensation (in US dollars) developing countries might seek from their fuel-guzzling companions on the surface of the Earth. Good sense suggests that it is too soon, in what is certain to be a protracted argument, to begin putting numbers to the damage that may

be done by global warming in such a literal fashion. This week's rumpus is more about the sense the technical critics of global warming have of their exclusion from the charmed circle of those with access to the most recent data from the climate models.

The pity is that IPCC is in poor shape to defend itself against such charges. Its three working parties deal respectively with the questions of climatic evidence, the consequences of warming and with "advice to policymakers". (The third group is that racked by the value of a human life.) The essential difficulty is that IPCC is both a quasi-judicial body, reviewing the published literature to see what sense it can make of it, and one that is meant to galvanize governments into action. Its publications are replete with "executive summaries" telling people what they must conclude even if they are too busy to follow the arguments. It is no wonder that IPCC is repeatedly in hot water. It would do better if it abandoned its second and third working groups for the time being, and if it persuaded its first working group to follow a more judicial course. That, sadly, is not likely to happen. □

China comes cleaner

China's bid to join the World Trade Organization does not mean that it is yet a modern state.

WHEN Groucho Marx said that he would not wish to belong to a club that allowed people like him to be members, it is unlikely that he was thinking of China. But China, formally still a Marxist state (in the other sense) is seeking membership of the capitalism's chief cheerleader, the World Trade Organisation (WTO) and the successor of the Geneva-based General Agreement on Tariffs and Trade (GATT). To strengthen its chance of being elected, the government of China announced earlier this week a series of new laws and regulations designed to make access easier. Tariffs will be reduced (from an average of 35 per cent to less than 20 per cent), the currency will be partially convertible and there will be simpler rules for telling how investments from abroad will be safeguarded. The obvious question is why China is taking all this trouble.

The simple answer is that China sees a huge competitive advantage, even by comparison with its neighbours in the Pacific. Even as things are, China's balance of trade with the rest of the world is roughly US\$20 billion a year in its favour; if its trading partners are prevented from discriminating against Chinese exports, it may expect to do even better. It may also expect the inward flow of investment (more than US\$40 billion in 1994) to grow as people elsewhere seek to take advantage of China's low labour costs. The other side of that coin is that membership of the WTO would place China more securely among the ranks of its trading partners and thus in the rest of the world. But a survey of "Science in China" to be published in *Nature* on 7 December will show that there is a long way to go before that is an unshakeable state of affairs. □

AIDS vaccine 'needs focused effort' as drug firms back off research

Paris. A world-wide consensus is emerging that the scientific community urgently needs to increase its efforts to develop an AIDS vaccine, partly to fill the gap left by the retreat of drug companies from pursuing this goal. Such a conclusion is being fuelled by the growing awareness that only a vaccine can curtail the explosion of AIDS in developing countries.

The need for a new, targeted vaccine strategy is, for example, expected to be the main thrust of a report on AIDS vaccine research at the US National Institutes of Health (NIH), to be released shortly by the agency's Office of AIDS Research (OAR) as part of a wider review by the office of AIDS research at the agency.

Broadly similar recommendations are thought likely to emerge from discussions of a 'task force' on vaccine research and development set up by the European commissioner for research, Edith Cresson.

The US report is widely expected to recommend that NIH should go beyond its traditional role of funding basic research and also become what Dani Bolognesi, from Duke University, the review's *rapporteur*, describes as a "discovery engine for the concepts for the design of the vaccine", aimed at bringing vaccines to the stage at which industry might regain interest.

"There is a sense that we have to do something radically different, with the major worry that otherwise companies will not become interested," says Bolognesi.

Moreover, while drug companies are traditionally reluctant to invest in any form of vaccine development — which carries high costs, low profits and big risks of costly legal suits should accidents occur — their current analysis of the state of AIDS vaccine research is particularly bleak.

"The traditional methods of vaccine

development don't work with HIV," says one spokesperson at Glaxo-Wellcome. "It's back to the [research] drawing board." Similarly, Maurice Hilleman, a director of the Merck Institute for Therapeutic Research in West Point, Pennsylvania, claims that companies are reluctant to invest money because "there is currently no basis for the development of an AIDS vaccine".

Some blame this reluctance on the way AIDS research has developed. "The field



Rising tide: an AIDS patient in Kigali, Rwanda.

has been driven by a discipline [molecular biology] and not by good science," claims one leading US vaccine researcher, arguing that this has led to the pursuit of approaches that ignore the pathogenesis of HIV.

He describes the current crop of subunit vaccines, such as those based on gp120 and gp160 antigens, as "the closest thing to insanity". Not only are they designed to reduce viral load — while HIV has high turnover and mutation rates — but they also suppress the cytotoxic T-cell response that clears infected cells and the virus they contain. The latter response is now considered to be an essential component of any effective vaccine.

The OAR review panel is said to have accepted such criticisms. As a result, the

panel is expected to recommend abandoning the approach of "simply pushing ahead and testing as you go along", according to William Paul, director of OAR. Instead, it is likely to propose a broader and more fundamental approach, building up a body of knowledge that would remain useful even if the next generation of vaccines should fail.

Similarly, the review is expected to urge NIH to set up a targeted programme. This would include preclinical studies to test vaccine concepts but would stop short of industrial development. Such research is not possible at present, says one vaccine researcher, claiming that funding research through investigator-initiated grants creates a "bunch of little islands" which are failing to accumulate the collective knowledge needed for industrial development.

Paul agrees that a more coordinated approach is needed. But he argues that this must not be at the expense of basic research on AIDS. Bolognesi also says that NIH needs to provide extra money for such programmes, in particular to improve the quality of much research by attracting leading scientists, and to make vaccine research a priority.

Bolognesi also points out that no systematic study has been carried out to determine which animal models are the most predictive of a good vaccine in humans. At present, models are not answering the right questions, he says; "we are mainly guessing".

Another AIDS researcher claims that the expected recommendations of the OAR review would represent a break with work over the past decade, during which AIDS vaccine research has been too distracted by, for example, the debate about whether to conduct clinical trials of subunit vaccines.

Similarly, Jean-Paul Levy, director of the French National Agency for AIDS Research, describes Phase III trials as an "absurdity" at the moment, given the lack of basic knowledge. "We are unsure if we will ever be able to get a vaccine for this virus, which is different to any virus for which we have ever created a vaccine," says Levy.

Not all are convinced that developing an AIDS vaccine means abandoning the traditional approach to vaccine development, however. "If we had been waiting to have all the immunological answers before we made the measles or polio vaccines, we would be still waiting," says Donald Francis of Genentech in San Francisco.

Francis argues that Genentech's gp120 candidate vaccine, which may enter trials in Thailand next year, has already protected chimpanzees and is safe, and that it is ►

Swiss role in EU research will remain limited

Basel. An agreement between Switzerland and the European Commission, which would have allowed Swiss scientists to play a more active role in the commission's Fourth Framework Programme, will not be signed this year because of the Swiss government's refusal to comply with the commission's demand that it open its labour market to all citizens of the European Union (EU).

The two issues are both part of a complex package of proposed agreements on relations between Switzerland and the EU that has been under negotiation since Switzerland abandoned plans to apply for membership of the European Economic

Area, as a prelude to EU membership, in 1992 (see *Nature* 373, 462; 1995).

Last week's deadlock in the negotiations means that Swiss scientists will not be able to apply for EC research money as project leaders from next year, as they had hoped. Under the present agreement, scientists can only take part in the framework projects led by non-Swiss scientists, and cannot become members of decision-making programme committees.

But Tim Guldemann, a member of the group that advises the Swiss government on science and research policy, says that he remains optimistic that agreement could be reached by 1997.

Oliver Klaffke

therefore worth testing. "The next-generation vaccines have not yet protected chimps," he adds. But one official at Genentech's parent company Hoffmann La-Roche says he is sure that "Genentech is well informed that gp120 is not very interesting for the future". Similarly, apart from the Thailand trials, Genentech's management is believed by many researchers to have pulled the plug on both the gp120 programme and AIDS vaccine research in general.

But most scientists seem to agree that it is time to explore a wider choice of research avenues. They argue, for example, that not enough attention has been given to so-called therapeutic vaccines which might attenuate infection to a level that could prevent the onset of AIDS, or transmission of the virus. Others argue that work on attenuated viruses merits much greater attention, despite the concerns about their safety.

In contrast, there has been growing criticism of the way that much AIDS vaccine research is directed at the B subtypes of HIV that are common in infections in the developed world. The OAR review is expected to recommend refocusing AIDS vaccine research on the HIV subtypes common in the developing world, where "a fire is raging", says one AIDS researcher.

The preliminary conclusions of the OAR reviews are echoed by an official at the European Commission, who claims that HIV vaccine research lacks both coordination and funding. The research directorate has recently created a task force on vaccines — headed by Bruno Hansen, formerly of Novo Nordisk — with other commission directorates. This will coordinate existing vaccine research within the European Union's Framework research programmes, and is seeking ECU100 million (US\$ 133.44 million) over the three years 1996–1998 for a series of new programmes.

Another proposal comes from Seth Berkley from the Rockefeller Foundation in New York, who is trying to raise \$600 million over seven years to set up an International AIDS Vaccine Initiative. This would fund existing laboratories rather than create new structures, says Berkley.

But others remain sceptical about whether the scientific community can replace industry's role in vaccine development. Levy, for example, argues that in the absence of adequate animal models, Phase I trials — which are needed not to test vaccines, he emphasizes, but to obtain information — require industrial support.

France is home to the company Pasteur Mérieux Sérums et Vaccins — now a subsidiary of Rhône-Poulenc Rorer — which is one of the few large companies taking a broad interest in HIV vaccine research. The company is already developing one of the first so-called second-generation 'prime boost' vaccines, made of an avipox virus genetically engineered to produce subunit proteins, and designed to give both a humoral and cellular response. **Declan Butler**

Black scientist faces inquiry at South African university

Cape Town. A black South African scientist with an international reputation in immunology, head-hunted from Britain to a senior post at the University of the Witwatersrand (Wits), has become the centre of a fierce controversy about his fitness for office, amid speculation over his possible candidacy as the university's next vice-chancellor.

The university has set up a tribunal to look into allegations by 13 senior academics against William Makgoba, one of its deputy vice-chancellors. In a letter to the current vice-chancellor, Robert Charlton, the academics challenged statements made by Makgoba in various versions of his *curriculum vitae*, and allege that he has failed adequately to carry out the responsibilities of his current job.

The academics also claim that a series of public statements made by Makgoba about the university over the past year have undermined — and are thus incompatible with — the proper performance of these responsibilities.

Makgoba in turn has dismissed the allegations against him as an "orchestrated campaign of vilification and disinformation". He has declined an invitation to discuss the matter with Charlton, arguing that the outcome of such a meeting would be a foregone conclusion, and asking that the matter be handled on a formal basis.

Makgoba graduated from the University of Natal Medical School in 1976, and left South Africa shortly afterwards to pursue his studies overseas. He appeared to have impeccable credentials for his new post at Wits, to which he was appointed last year; at the time of his appointment, for example, he was a senior lecturer in immunology at the Royal London Postgraduate Medical School in the United Kingdom.

But last month, scarcely a year after he assumed office, a group of nine deans and four senior professors at Wits wrote a letter of complaint to Charlton. The allegations against Makgoba include falsely claiming membership of both the American Association of Immunology and the British Transplantation Society, and receiving a grant of £80,000 (US\$128,000) from the Nuffield Trust while at Oxford's John Radcliffe School of Medicine between 1979 and 1983.

Makgoba also wrongly claimed that a paper he co-authored in the journal *Immunology Today* in 1988 was the most-

cited article in the life sciences in the following year. In the light of these charges, Charlton sent a letter inviting Makgoba and his legal representative to a meeting on 2 November, saying that this meeting was intended to satisfy him that "the allegations lacked weight and substance, and that there would be no need for further proceedings".

But Makgoba, who could not be contacted for comment, has said that he prefers to clear his name in front of a full committee of inquiry. "The thirteen excellent academics got it so wrong it is incredible," he told the *Johannesburg Sunday Times*. "In conducting research by using Procrustean methods they will follow the fate of Procrustes," he said, adding that the Greek mythological figure — who used force to impose conformity — was himself eventually killed.

Some of Makgoba's critics have themselves come under fire for their tactics. Charles van Onselen, for example, a social historian, has been criticized for the methods he used to investigate Makgoba's claim that he was the first African graduate to obtain a distinction and a certificate of merit in medicine at the University of Natal. Rather than requesting the information directly, Onselen learned that no such award had been made by writing to the university registrar on the pretext of obtaining assistance with what he called "a rather esoteric inquiry for a piece of social history".

Last week, representatives of the Wits Black Staff Forum and the South African National Students' Congress accused a "white clique" of waging a vendetta against Makgoba and called for Charlton's resignation on the grounds that his handling of the matter and statements have polarized the university on racial lines. Makaziwe Mandela, an executive member of the forum, and daughter of President Nelson Mandela, claimed the university was trying to get rid of Makgoba because he had demonstrated his independence and intelligence.

But Charlton, who retires in 1997, has been trying to play down the racial dimension. He has responded to critics such as Mandela by claiming that, while everything in South Africa is represented in racial terms, the complaints against Makgoba can be investigated objectively. "If he had been white, the complaints would have been the same," he added.

A three-person tribunal has been set up by the university to look into the affair. Lord Flowers, the former vice-chancellor of the University of London, and Walter Kamba, dean of law at the University of Namibia, have already agreed to be members. The third person has yet to be appointed.

Michael Cherry



Makgoba: tribunal will investigate allegations.

US space agency escapes the worst of cuts

Washington. In the midst of gridlock and government shutdown, the US Senate and House of Representatives quietly settled their differences last week over funding levels for three key science agencies: the National Aeronautics and Space Administration (NASA), the National Science Foundation (NSF) and the Environmental Protection Agency (EPA).

NASA's Mission to Planet Earth, a major programme of climate and environmental monitoring, emerged as the least unscathed, with a cut of only \$81 million to the Clinton administration's original request of \$1.34 billion — \$20 million less than the Senate had approved, but \$257 million more than the House had granted.

The NASA vote was a defeat for House Republicans, particularly Robert Walker (Republican, Pennsylvania), chair of the House Science Committee, who had sought

deep cuts in NASA's Earth Observing System (EOS) of satellites. In contrast, it was a victory for Senator Barbara Mikulski (Democrat, Maryland), whose state is home to the Goddard Space Flight Center, where much of the EOS work will be carried out.

In preserving NASA's Earth science programme, the conference members may have been influenced by a recent National Research Council study that argued strongly against cutting near-term EOS funding, while suggesting that future savings could come from redesigning a proposed system for distributing data to scientists (see *Nature* 377, 191; 1995).

Other threatened NASA projects also fared well in the conference. The proposed Space Infrared Telescope Facility (SIRTF), for example, lost only \$5 million of a \$15-million request, rather than being eliminated as the House had proposed. That

should allow the project to continue plans for a launch in 2002.

Legislators participating in the funding 'conference' neatly split the \$40-million difference between proposed House and Senate funding levels for the NSF, setting the final mark at \$3.18 billion. Under the compromise plan, the 'research and related activities' account, which supports most NSF grants, increases slightly over 1995 levels.

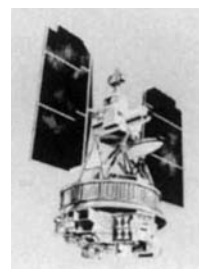
EPA fared better in the compromise budget than it had in either chamber. The conference allocated \$5.7 billion for the environmental agency, an increase of \$48 million over the Senate mark and \$817 million over the House allocation. That still may not be enough, however, to prevent a veto by President Bill Clinton. He wants more money both for EPA and for his proposed programme of national service, which, along with federal housing programmes, is funded by the same bill.

Arguments about the legislation were expected to continue at least into this week. But negotiations from now on are unlikely to involve NSF and NASA, so the levels set for them last week are likely to stand.

Meanwhile, workers at those two agencies and EPA who had been temporarily laid off returned to work this week after Clinton and Congressional Republicans came to an agreement over the weekend on general principles for balancing the federal budget. A new continuing resolution provides temporary funding (see left) until mid-December, or until the appropriations bill is passed.

During the four-day government shutdown, some scientists managed to get into their offices to keep up with workloads. But most could not, and telephones at all three agencies rang unanswered. Not all of NASA's civil servants were laid off. For example, personnel at Ames Research Center responsible for monitoring the Galileo probe, which arrives at Jupiter on 7 December, were among those considered 'essential' and allowed to report for work. But a Galileo pre-arrival press conference and an educational session for 300 teachers were both cancelled.

Even senior NASA officials felt the effects of the shutdown. When Daniel Goldin, the agency's administrator, wanted to make a congratulatory call to shuttle astronauts meeting their Russian counterparts in orbit, he found his headquarters office closed. He ended up telephoning from his living room.



Early bird: TOMS ozone mapper is seeking successors.

NIH discovers who counts as 'essential'

Washington. Last week's unprecedented six-day shutdown of the US government halted the work of most grant-giving agencies, and sent home thousands of scientists and support staff at government laboratories.

At the main campus of the National Institutes of Health (NIH) at Bethesda, Maryland, only those scientists, administrators and other employees considered 'essential' were permitted to work — numbering about 2,000 of the 16,000 NIH staff. The library was closed, seminars were cancelled, telephones were left unanswered, and the ordering of supplies was at a standstill.

Most centres of the National Aeronautics and Space Agency (NASA) were also closed (see above) and all but 22 of the National Science Foundation's 1,300 staff were sent home. But the Department of Energy's laboratory network worked as normal, as the department is one of a handful for which a budget had been agreed.

"We have just about run out of [the radioactive isotope] P-32," said Robert Nussbaum, chief of the laboratory of genetic disease research at NIH. "This is not fun. [Our research is] being seriously interfered with." Nussbaum and others said they would have had to wind down experiments if the shutdown had continued for longer than two weeks.

The closure, however, ended on Monday, after the Clinton administration and Congress agreed to a temporary spending measure that will fund programmes until 13 December. Under a complex formula, the measure will fund NIH at its

1995 level until then, and most other agencies at the lower level proposed by the Senate or the House.

NIH laboratories working with cell lines, animals and patients were allowed sufficient personnel to avoid any damage or harm. But the definition of 'essential' workers was interpreted differently by individual laboratories — resulting in widespread resentment among some of those laid off. NIH stuck to its hierarchical structure in making lay-offs: laboratory heads were asked to choose who would stay and who would go. Final decisions were made by the intramural programme office.

"It was a psychological blow to those considered non-essential," says Vincent Hearing, a biochemist with 25 years experience at the National Cancer Institute. Most laboratory chiefs appear to have been deemed essential, and were at work last week. But Clinton and Congress agreed that those who were sent home will still be paid. A spokesperson for the NSF — speaking from home — said that Neal Lane, the NSF director, decided to lay off everyone except himself, his deputy and security staff, in order to minimize "morale problems" if some staff were considered 'essential' and others not.

James Sheirer, deputy director of the division of extramural affairs at NIH, confirmed that his office, responsible for administering grants, was "at a dead stop". NIH awards to be made on 1 December may be delayed, as will most other grant-related business.

Adrianne Appel & Colin Macilwain

Biosafety rules will regulate international GMO transfers

London. Signatories to the United Nations Convention on Biological Diversity, which concluded their second annual conference in Jakarta last week, have agreed to formulate an international protocol regulating the transfer of genetically modified organisms (GMOs) between states.

The agreement represents a compromise between the Group of 77 (G77) countries, which favoured early enactment of a comprehensive protocol on biosafety, and the industrialized nations, which have been seeking the adoption of looser guidelines.

A group of experts has been given the task of working out details of the protocol, which will not come before the convention's signatory nations until at least 1999. The protocol will regulate accidental releases of modified organisms in transit. But it will not extend to regulating the handling and domestic use of GMOs, a goal enshrined in the biodiversity convention, signed at the Rio Summit in 1992.

Last week's agreement is largely the work of environment ministers from the Euro-

pean Union (EU), spearheaded by Britain. The present terms were agreed by EU ministers at a meeting in Luxembourg in July. EU delegates to the biodiversity conference then successfully persuaded G77 waverers at the Jakarta meeting to support the new alternative. Reaction to the agreement, reached on the conference's penultimate day, has generally been favourable.

test and release "potentially dangerous and self-replicating organisms in countries with weak or non-existent biosafety laws".

The EU maintains that domestic laws on biosafety should be the responsibility of national governments, not of the UN. Calestous Juma, the new executive secretary of the biodiversity convention, argues that such countries can now rely on UN biosafety guidelines — to be finalized next month — as a basis for national legislation.

Apart from discussing biosafety, delegates representing 117 of the convention's 134 signatories passed a mandate urging governments to conserve marine and coastal biodiversity. The United Nations Environment Programme also used the conference's second day to launch its long-awaited 1,000-page inventory of the Earth's biodiversity.

But several issues remain pending. For example, a decision on the terms under which companies should be allowed access to the genetic resources of other countries, and a debate on the patenting of indigenous innovations, were both put off to next year's meeting in Argentina.

Similarly, whereas the Jakarta conference agreed on a two-year pilot phase for the so-called 'clearing house' on scientific and technological cooperation, developed countries are thought to be cautious about the terms and nature of the information to be made available, which have yet to be discussed.

One of the Jakarta meeting's more controversial resolutions was an invitation to the World Trade Organization (WTO) to prepare a report on the relationship between

the biodiversity convention and the WTO's Trade Related Intellectual Property Rights Agreement (TRIPs).

Friends of the Earth points out that such an invitation represents a conflict of interest. The biodiversity convention grants sovereign rights over genetic resources to farmers and indigenous peoples; but under TRIPs, such resources, once genetically modified, become the property of the multinationals that carried out the work. "The WTO is simply not in a position to carry out an unbiased assessment of the agreements," says Ronnie Hall. An assessment needs to be carried out by independent experts, she adds.

One move which appears to have paid off is the inclusion of non-governmental organizations such as Greenpeace and Friends of the Earth inside closed contact-group negotiations. NGOs are usually kept at arm's length of such meetings, such as in the working groups of the Intergovernmental Panel on Climate Change (IPCC). **Ehsan Masood**

Swiss researchers adopt voluntary rules on exports

London. Swiss research institutions, keen to counter criticism of the lack of rules covering the export of genetically manipulated organisms (GMOs) to researchers in other countries, have agreed to adopt a set of voluntary safety guidelines covering such exports.

The new guidelines have been drawn up by the Swiss Interdisciplinary Committee for Biosafety in Research and Technology (SCBS), and are expected eventually to be given legal backing. They include the requirement on an exporting laboratory, whether public or private, to notify an appropriate agency in the recipient's country of the nature and quantity of the organisms involved, and to provide the agency with a brief summary of "the assessment of risk to human health and the environment".

The change has been prompted partly by a widely publicized incident last year, when the environmentalist group Greenpeace intercepted a package of genetically modified rice seeds being sent to the International Rice Research Institute in the Philippines.

"One reason for this Greenpeace action was to alert the public to the fact that no legally binding regulations exist in Switzerland, or internationally, covering the exports of GMOs to be used in field trials," says Karoline Dorsch-Häslar, secretary to the committee. The public dispute took place, she points out, despite the fact that the Swiss researchers had obtained permission from the Philippine government to import the rice.

Under the new guidelines, which will allow the committee to keep track of all GMO exports from Switzerland, any exporter has to obtain the prior "informed agreement" of any country to which GMOs are sent. The exporting organization also has a responsibility to comply with the appropriate safety procedures in the receiving country.

The guidelines are based on procedures that have been drawn up by the United Kingdom and the Netherlands, which were recently applied for the first time to a bilateral agreement between Britain and Argentina (see *Nature* **377**, 97; 1995).

Although they do not yet have the force of law, the guidelines have already been accepted by most of the bodies responsible for publicly funded research in Switzerland, as well as by the Society of Swiss Chemical Industries.

David Dickson



Open for business? Gene bank at Sebele, Botswana.

pean Union (EU), spearheaded by Britain. The present terms were agreed by EU ministers at a meeting in Luxembourg in July. EU delegates to the biodiversity conference then successfully persuaded G77 waverers at the Jakarta meeting to support the new alternative. Reaction to the agreement, reached on the conference's penultimate day, has generally been favourable.

Andrew Dickson, a spokesman for the International Bioindustry Forum, describes the protocol as a relevant, welcome and sensible conclusion and expresses satisfaction that "industry's voice has been heard more strongly than hitherto". Nancy Vallejo, coordinator of international treaties at the World Wide Fund for Nature, said that agreement on a protocol was "a significant result".

But not everyone has been persuaded. Ronnie Hall, for example, of Friends of the Earth, says legislation on the domestic handling and use of GMOs is still needed. Without such measures, she adds, companies can

Guidelines may boost gene therapy in Japan

Tokyo. Gene therapy in Japan took a further step forward this week with the publication of guidelines for the manufacture and testing of therapeutic products by the Ministry of Health and Welfare (MHW) that are expected to pave the way for greater activity.

The new guidelines clarify the vectors, methods of administration, testing and certification needed for approval of products used in gene therapy, and are virtually identical to those already published by the US Food and Drug Administration (FDA).

Gene therapy has been slow to take off in Japan. Only one clinical trial has begun so far, and the pharmaceutical companies — none particularly large by world standards — have been slow to enter the field.

According to Yoshiro Ohtaki of Japan Associated Finance Corporation (JAFCO), a venture capital company that is backing various biotechnology initiatives, this is partly due to the lack of qualified personnel for controlling vector quality and other tasks specific to gene therapy. But another factor has been the lack of official guidelines for companies to follow in the formulation of therapeutics, with the result that companies have not known what data was required.

The delays experienced by Japan's first gene therapy trial, which started in August at Hokkaido University in northern Japan (see *Nature* 376, 628; 1995), were caused by the fact that the health ministry had not initially been supplied with safety data on the viral vectors by the manufacturers, Genetic Therapy Inc., of Gaithersburg, Maryland.

The Hokkaido trial, which has involved treatment of a four-year-old boy suffering from adenosine deaminase (ADA) deficiency, is proceeding successfully. It will not lead

to a commercial therapy, as there are relatively few sufferers in Japan. But its success has helped the broader prospects for commercial gene therapy in Japan by demonstrating the efficacy of the treatment to a sceptical pharmaceutical industry, according to observers such as Mitsuru Miyata, chief editor of *Nikkei Biotechnology*.

Japan's first clinical trials of a potentially commercial gene therapy will probably begin early next year when Midori Juji (Green Cross Corporation), an Osaka-based manufacturer of blood products, will begin testing a treatment for HIV infection using viral vectors supplied by the US company Chiron-Viagene.

The Japanese company, which will carry out its research in association with Kumamoto University Hospital in Kyushu, has informally told both the health ministry and the Ministry of Education, Sport, Science and Culture (Monbusho) of its plans. But it has yet to make an official

application for approval for the research.

Although the new guidelines will hasten the application of gene therapy technology in Japan, formidable barriers remain to its indigenous development. Of the five groups planning clinical trials of such therapy at universities in Nagoya, Nigatta and Tokyo, four will rely on vectors supplied by US companies, says Kenzaburo Tani of Tokyo University's Institute of Medical Science.

Tani, whose group plans to test a gene therapy for renal cell cancer, says that although Japan's pharmaceutical companies are reluctant to pay the royalties demanded for the use of foreign technology, they are also reluctant to support the basic research required to develop their own technologies. In contrast to the United States, where the need for high risk research is met by start-up companies financed by venture capital, the continuing lack of such companies in Japan appears to be leaving its own industry at a disadvantage.

Stephen Barker

Review for radiation effects research

London. The United States and Japan have agreed to set up an international panel to review the current and future activities of the Radiation Effects Research Foundation (RERF), the body set up in Hiroshima to study the long-term effects of the atomic bombs dropped on Hiroshima and Nagasaki at the end of the Second World War.

The panel will be chaired by Roger Clarke, chairman of the International Commission on Radiological Protection and director of the UK National Radiological Protection Board (NRPB). Its eight mem-

bers — none of whom have previously been involved in RERF's activities — include two medical researchers from Japan, two from the United States, and one each from Argentina, Australia, Germany and the United Kingdom.

The decision to set up a review panel follows the protests that greeted plans announced by the US Department of Energy (DoE) at the beginning of the year to remove administrative and scientific responsibility for the RERF from the National Academy of Sciences, which had carried out this role since 1947. Apparently backtracking on earlier claims that the management contract was being withdrawn from the academy partly because of excessive secrecy surrounding some of RERF's research, Hazel O'Leary, Secretary for Energy, subsequently announced that the academy's contract was to be extended for a further two years (see *Nature* 375, 709; 1995).

At the same time, the department said that it had accepted a proposal from RERF's Science Council that any restructuring of the foundation's work should be deferred until it had been reviewed by a team of external experts. The panel that has now been set up is expected to identify how the foundation can use its data and expertise to contribute to radiation research worldwide.

The first meeting of the panel, whose broad remit is "to strengthen and enhance the scientific contributions of RERF", will take place in February, and will include visits to both Hiroshima and Nagasaki. Technical support will be provided by the NRPB, and a final report is due by July 1996. □

Livermore loses supercomputers to Berkeley

San Francisco. Lawrence Livermore National Laboratory (LLNL) in California has lost its \$45-million supercomputing centre to its northerly neighbour, Lawrence Berkeley Laboratory, after a competitive bid. The National Energy Research Supercomputer Center (NERSC), housed at Livermore for 22 years, employs 110 people in unclassified energy research and manages three supercomputers, including a \$25-million Cray C-90.

US laboratories have been under considerable pressure to reduce spending and streamline their activities. They have been asked to find more opportunities to cooperate on projects and share resources. Livermore scientists will continue to use the NERSC through high-speed links.

The Department of Energy (DoE) said the laboratory's location at Berkeley will allow integration of non-classified computing with research programmes in energy science, chemistry, high-energy and nuclear

physics, material science and genomics. It plans to complete the move by next April.

Berkeley's proposal had convinced DoE that it could manage the computer laboratory more efficiently and with more potential for collaboration. Both laboratories aimed to cut costs by about 20 per cent.

Staff members at Livermore expressed their disappointment with a "for rent" sign on the door of the centre. But Bruce Tarter, director of LLNL, played down the importance of the supercomputers' departure. He has long argued that supercomputing can still grow at the laboratory, even without NERSC. A new LLNL team, made up of people from Livermore Computing (LC) and NERSC, will reorganize computing for both classified and unclassified research at the laboratory.

NERSC was created in 1974 to serve the national magnetic fusion energy programme. According to LLNL, it currently serves about 4,500 users. Sally Lehrman

Europe tries again on biotechnology patents

Munich. As the European Patent Office (EPO) prepares to rule this week on a challenge to the patentability of Harvard University's 'oncomouse', the European Parliament has agreed to take part in a three-way discussion next January with representatives of the European Commission and the Council of Ministers on various issues relating to biotechnology, including new rules to harmonize patent legislation.

The draft rules are being drawn up by the commission, following the unexpected rejection of its previous proposals by the European Parliament last March, after they had been approved by the European Council of Ministers (see *Nature* 374, 103; 1995).

A broad political spectrum of parliamentarians joined their Green colleagues in rejecting the bill. In addition to the Greens' opposition to all patenting of any life forms on moral grounds, two other factors were cited. First, the draft directive did not protect the rights of farmers to breed from their own animals for their own purposes, known as 'farmers' privilege'. Second, many claimed that a lack of clarity in the text left too many loopholes through which individual member states could squeeze.

Christof Tannert, for example, a German

member of the European Parliament (MEP), and speaker for the European Social Democrats on biotechnology, claims that the proposed text did not distinguish sufficiently clearly between a discovery and an invention. This would have allowed for a sort of 'biological colonialism', whereby industries in developed countries could patent newly discovered species of plants or animals in developing countries, he says. In addition, he expressed concern that the patenting of techniques for germ-line therapy was not explicitly excluded.

Commission officials admit that the draft text, the result of seven years of negotiation with both member states and the parliament, still contained some ambiguities, and has prepared a new version. This time, it has launched a broad campaign to increase the chances of its approval.

Over the past few months, for example, Mario Monti, commissioner for the internal market and industrial affairs, has consulted widely among parliamentarians and industry about whether the new text — which is not yet publicly available — will be acceptable. January's discussion meeting will be a final test of mood before the commission announces a formal decision to put the new

directive on the table.

Dominique Vandergheynst, the EC official who is the chief architect of the revised draft directive, says that the new text acknowledges a 'farmers' privilege', specifically excludes patents on germ-line gene therapy techniques, and distinguishes clearly between invention and discovery, with the latter excluded from patentability.

With these changes, Vandergheynst claims that most MEPs should find the new directive acceptable, predicting that its first parliamentary reading should be completed within a year. According to Vandergheynst the main thrust of what many considered the most controversial section — the one that covers the patenting of human genes — remains unchanged from the earlier draft.

Thus genes in their natural state would be unpatentable. But isolated genes, when used diagnostically or as starting material for a biotechnological process, would be patentable. Transgenic animals would in principle be patentable, provided that their use for human well-being outweighs any suffering to the animal, as would plants, but not plant varieties.

The biotechnology industry, which has cooled on the need for a directive following the debacle in March, is now waiting with interest to see the new text in detail. Brian Yorke, head of the patent department at Sandoz, says that a good directive is needed to harmonize national patent laws, and to give a clear ruling on whether genes, plants and animals are patentable, as this remains ambiguous in the European Patent Convention. "Industry must know what the legal position on infringement in Europe is in this field," he says. He claims that the original, unsuccessful draft, was too unclear to be of significant benefit to industry.

Yorke says that he is very optimistic about the revision. But, like many in industry, he is concerned that further changes could be introduced during the lengthy 'co-decision' process under which all three European bodies — commission, council and parliament — must agree on all directives.

A spokesman for the European Federation of Pharmaceutical Industries puts it plainly: parliamentarians, who have considerable influence as a result of the co-decision process, may now be reassured, but are very unpredictable and may come up with new problems.

Furthermore, not all of industry is enthusiastic about a new directive. Jan Leemings, director of Plant Genetics Systems (PGS), argues that rulings on patent protection should be left to the EPO. He says that the EPO is already resolving the lack of clarity in the European Patent Convention by building up case law experience, and is concerned that a further layer of legislation could add new restrictions. **Alison Abbott**

Agronomist quits in funding protest

Munich. A prominent French agronomist has resigned angrily from the committee responsible for a European programme in agriculture-related research in protest at what he claims to be excessive political interference in the allocation of funds.

Roger Cassini, director of the Paris-based Institut Nationale de la Recherche Agronomique, has resigned as French delegate to the programme committee of the European Commission's ECU600-million (US\$800-million) programme on fisheries and agro-industry research (FAIR), accusing the commission of having manipulated a list of successful grant applicants to ensure its approval by EU member states.

The commission itself has so far refused to comment on Cassini's resignation, which was triggered by a dispute over the distribution of ECU44 million which the commission had available for allocation to applicants responding to its first call for proposals for the FAIR programme, part of the fourth Framework programme.

Having sifted through nearly a thousand applications, almost half of which were judged by scientific reviewers as suitable for funding, an initial shortlist was drawn up by the commission's staff last month. But this was rejected by the programme committee, whose members are representatives of EU states, some of whom — particularly Italy, Spain and Greece — did not feel that

their own scientists were fairly represented among the shortlisted applicants.

Faced with the prospect of losing the money unless an approved shortlist was adopted by the commission by the end of the year, the commission presented a revised shortlist to the programme committee last week. But it withdrew plans to put the list to the vote when it became clear that it was likely to be rejected again.

According to Cassini, the director of the committee then left the room with the delegates from Spain, Italy and Greece. He returned shortly afterwards with a revised list, with two additional projects added favouring these countries at positions which virtually guarantee funding. The three countries then announced their support for the revised list.

Cassini describes the commission's actions as "scandalous". A British delegate to the committee describes the meeting as chaotic, and claims that the dispute over the shortlist has highlighted the controversial role of national interests in allocating research funds. The commission does not have rules requiring it to achieve geographical balance in its programmes.

But another committee member, Christian Hubrich, of Germany's federal ministry of agriculture, says that at the same time the revision of the shortlist, while regrettable, did not infringe any rules. **A. A.**

Pakistan's scientists respond warily to World Bank reforms

London. The Pakistan government has been advised by the World Bank to carry out a broad-ranging shake-up of its research institutions in a belated attempt to harness science to wealth creation.

The changes, if carried through, would represent the most far-reaching upheaval since reforms in the late 1950s. The centrepiece would be the closure of Pakistan's oldest and largest multidisciplinary research organization, the 43-year-old Pakistan Council for Scientific and Industrial Research (PCSIR), which lists the development of Slimfast, an effective weight-reducing diet, as its key achievement of 1994.

But the proposals have caught unawares the country's scientific community. While giving a cautious welcome to the proposed changes, some scientists believe the government appears to be abdicating its responsibility for promoting science.

While the government insists it is still considering the proposals, an *aide-memoire* from a recent World Bank mission to Pakistan in July suggests that changes are well advanced. Their main emphasis is on restructuring the PCSIR's nine laboratories into efficient, single-discipline institutions, focusing on the strategic areas of research in glass, biotechnology, applied chemistry, fuel technology, pharmaceuticals and minerals.

Each institute would be managed by a chief executive officer and guided by a board of directors. They would be presided over by a newly created Pakistan Research and Technology Foundation whose principal aims would be technology "foresight and R&D project management". Of the estimated cost of US\$80 million, \$18 million would come from the private sector, \$44 million from World Bank loans, and the remainder from the government.

Basit Hasan, a spokesman for the Pakistan Association for Scientists and Scientific Professionals (PASSP), says many researchers in Pakistan were not even informed of the World Bank's proposals.

Many believe they are being blamed for PCSIR's failure to turn science into commercial applications. But Hasan says this is unfair, as the government, still lacking a minister for science, has made little effort recently to bring industry closer to research. The PCSIR, he adds, had been expected to develop world-beating technologies on a budget of less than \$10 million a year.

Javed Jabbar, a former minister for science and architect of the National Technology Policy, says that the World Bank proposal that the private sector should take on the main responsibility for bailing out science is unworkable in Pakistan.

Ehsan Masood

Climate critics claim access blocked to unpublished data

Washington. Researchers who do not believe human activity is having a significant impact on global climate change have been denied access to the latest data from British climate models used by the Intergovernmental Panel on Climate Change (IPCC), a congressional hearing was told last week.

The allegation was made by Patrick Michaels of the University of Virginia, an outspoken sceptic about man-made climate change. But the group that carried out the research has justified not making the data available on the grounds that it had not yet been published in the scientific literature.

Michaels was speaking at a hearing of the House of Representatives Science Committee's energy and environment subcommittee, chaired by Dana Rohrabacher



(Republican, California). The hearing was the second of three on "scientific integrity and public trust".

According to Rohrabacher, the purpose of the hearing was to ensure "that both sides are presented" on the climate change issue. But Democrats and environmental pressure groups say that the hearing, like an earlier one on ozone depletion, merely gave excessive prominence to the views of a noisy handful of "dissident" scientists.

Michaels told the subcommittee that "taxpayer-supported organizations" were denying "critical scientists" access to data required quantitatively to review new syntheses on global warming. He said he had asked John Mitchell of the UK Meteorological Office on 10 May for detailed output data from an improved climate model then being used by the IPCC in preparation of its Second Scientific Assessment.

Michaels wanted data from the model — which had been modified to take account of the cooling effects of sulphate aerosols — to enable him to test his theory that the improved model would still predict too much warming at high latitudes.

Mitchell declined the request on the grounds that the data had not yet been published; the paper appeared in *Nature* in the issue of 10 August (376, 501; 1995). But Michaels, speaking after the hearing, argued that he should have been given access to the data earlier on the ground that the work had been accepted for publication and was, in any case, already being used by the IPCC, whose work he was seeking to review.

Tim Roemer (Democrat, Indiana) told Michaels that his "problems with the UK" were beyond the jurisdiction of the US Congress, but Michaels countered that the work was funded by United States taxpayers. In fact, the Mitchell work was funded by the UK Department of the Environment, but the IPCC is a United Nations organization, and, therefore, would be funded by US taxpayers if the US paid its dues to the UN.

Disputing the theory of global warming caused by human activity, Michaels told the hearing that he expected "a change of paradigm towards the idea that the Earth is a great deal more resilient than we thought". But Jerry Mahlman, director of the Geophysical Fluid Dynamics Laboratory at Princeton, New Jersey, replied that anthropogenic global warming is "a harsh and inexorable reality".

In addition to their disagreement on the science, Roemer and Rohrabacher also clashed on budget matters, with the former claiming that the hearing was being held too late, as congressional action was virtually complete on global change research.

A key feature of the series of hearings has been various challenges to the integrity of the scientific basis to claims about contemporary environmental threats. But Lynn Rivers (Democrat, Michigan) turned the tables by calling into question the credibility of witnesses making such allegations.

In particular, Rivers said that enquiries by subcommittee staff had failed to substantiate a charge by Sally Baliunas of Harvard University that she had lost funding because of her views on ozone depletion. "Despite the allegations by Drs Baliunas and Michaels, the subcommittee has received exactly zero evidence that there is some sort of cabal among environmental scientists to keep these critics unfunded and voiceless," she said afterwards.

Rivers, who tried unsuccessfully to reintroduce the issue of Baliunas's allegations into last week's hearing, added that these critics "are not involved in scientific debates and discourse because they find it easier to play to the popular press and to seek funding from groups whose interests align with these scientists' predispositions".

Colin Macilwain

Atomic energy in Pakistan

SIR — I wish Ehsan Masood had been more objective in his report on the Pakistan Atomic Energy Commission (PAEC) (*Nature* 376, 633; 1995). In contrast with his discussion of the Pakistan Agricultural Research Council, there is no mention of how many papers have been produced by PAEC scientists, or the judgement their work receives from their colleagues through citations. Instead, Masood quotes the description by the director of PAEC's Institute for Nuclear Science and Technology (PINSTECH) of his staff as "the largest collection of trained, motivated and research oriented scientists in the country".

It is not that the publication figures are hard to find. A quick glance at the PINSTECH Silver Jubilee Technical Report shows that its staff has managed to produce only 679 papers in international journals in the past 25 years. This amounts to barely 25 papers a year, from the "largest collection of trained... scientists" and is hardly the kind of performance that justifies the label of "jewel in the crown" that the author confers.

On the other hand, the author dwells on the technological achievements of the PAEC and PINSTECH, claiming that PINSTECH is "in the midst of building" a 300-MW reactor at Chashma, "with the help of China". The implied division of labour is clear: PAEC is doing the building and the Chinese are lending a helping hand. Only the PAEC would describe it thus.

Independent sources such as *Nuclear News* (February 1992) described the deal as one where "China will provide a 300 MW pressurised water reactor to Pakistan". The reactor, which *Nuclear News* describes as an "import", is based on Qinshan-1, which China considers to be an "indigenous design". *The Far Eastern Economic Review* (23 January 1992) in its report puts it as follows: "An agreement for China to build a nuclear power-station in Pakistan." The details it gives are that the Chinese are to "install" the reactor, "supply nuclear fuel to run it" and "transfer nuclear technology to Pakistan".

In fact, the Chinese are selling Pakistan a largely untested Chinese-designed reactor for a "give-away" price of US\$500 million (according to *The Far Eastern Economic Review*) as part of their plan to get into the nuclear supply business. There is no way this could be described as the PAEC being "in the midst" of building anything except a cosy little empire, one in which there is no financial accountability, no independent scientific oversight of the programmes, and contempt for parliament.

The reason that the PAEC is allowed to get away with this is simple. It is hard for

many in Pakistan to challenge the PAEC because of the link with the nuclear weapons programme, with the PAEC only too happy to wrap itself in the cloak of national security. Your report, incidentally, omits mention of the secret 40-MW plutonium production reactor that Pakistan has been trying to build at Khushab, a project that came to light during the visit of the prime minister of Pakistan, Benazir Bhutto, to the United States earlier this year (*Washington Post* 8 April 1995).

Zia Mian

*Sustainable Development Policy Institute,
PO Box 2342,
Islamabad, Pakistan*

SIR — Ehsan Masood's report on "Science in Pakistan" (*Nature* 376, 631–638; 1995) deserves a more extensive response than can be published as a simple letter. I shall therefore limit myself to a single point—the misrepresentation of the role of Professor Abdus Salam, a Nobel prizewinner, in Pakistani science.

The writer makes three claims. He would have us believe that Salam pushed particle and nuclear physics at the expense of more "practical" sciences, that he succeeded in foisting these views on Pakistani universities to their detriment, and that he voluntarily exiled himself from the country of his birth.

In fact, Salam rarely missed an opportunity to urge Pakistani physicists and graduate students to work in practical areas of physics such as condensed matter, laser, plasma and so on. Coming as it did from a Nobel prizewinner in high-energy physics, I found this view a little unsettling at times because high-energy physics is my field and I do believe it to be important. But Salam was firm on this point in a discussion that I had with him almost a decade ago. Pakistan needs technology right now, he said, so let's leave theoretical physics to a small number of truly gifted people.

Masood goes on to credit, or rather impugn, Salam with "guiding the graduate physics programme at Quaid-i-Azam University" (Pakistan's leading university) and claims that "Pakistan's university physics departments are a monument to Salam's influence". No one would wish for such an awful monument; if Salam had indeed influenced our physics departments, maybe they would not now be so intellectually impoverished. Far from guiding or founding our physics programme, Salam has never even visited Quaid-i-Azam University, where I have been teaching since 1973.

I would like to make it clear that this has not been by choice on his part, or on the part of the faculty. After Salam received the Nobel prize in 1979, several of the physics faculty wanted to invite him

to a symposium at the university. But when the Islami-Jamiat-i-Ealaha, the student wing of a fundamentalist political party, threatened to disrupt the meeting violently if he was invited, the invitation was withdrawn. This group and its parent party consider Salam to be a heretic because he belongs to the Ahmadiyya sect.

This relatively minor episode brings me to my third point. Masood reports the allegations of Salam's detractors that he abandoned his native country for the West. Having worked as a physicist in Pakistan for more than 20 years, it is my considered opinion that Salam's talents would have rapidly atrophied in the intellectual desert of Pakistani universities. Had he stayed on in Pakistan, Salam could never have blossomed into a leader of the most difficult and competitive field of scientific endeavour. Nor could he have founded the International Centre for Theoretical Physics in Trieste, Italy, an invaluable source of help to science development in many countries. As a resident Pakistani, Salam would have had negligible influence in either science policy or institution making.

There is a still graver issue: given the violence regularly and systematically perpetrated against the members of his religious community, it is far from clear that Salam could have survived even physically. A year ago, on the night of 9 October 1994, I rushed the bullet-torn body of my departmental colleague and next-door neighbour, Dr Naseem Baber, to a hospital in Islamabad. He too was a member of the Ahmadiyya sect.

Pervez Hoodbhoy

*Physics Department,
Quaid-i-Azam University,
Islamabad 45320,
Pakistan*

Residents' rights

SIR — In response to my earlier letter, Michel Cuénod, the secretary general of the Human Frontier Science Program (HFSP), points out (*Nature* 377, 192; 1995) that nationals of any country can apply for HFSP funding as "co-applicant" as long as the principal applicant is a citizen of one of the member states.

But there is an important principle at stake here. In spite of their lack of citizenship, permanent residents are, nevertheless, allowed to pay taxes. HFSP member states use these tax revenues to contribute to the programme. So should not permanent residents enjoy the same access to HFSP funding as citizens?

Michael Lardelli

*Department of Developmental
Neuroscience,
Uppsala University,
Box 587, BMC,
751 23 Uppsala, Sweden*

Patterns of pattern formation

As participants at *Nature's* conference "Patterns of Life" (Boston, 9–10 November) found, moves to different organisms and even to biochemistry are helping to sustain the drive to understand how different body plans develop.

LAURENCE Sterne, being an avid reader of Ephraim Chamber's original *Cyclopaedia*, was well acquainted with the contemporary scientific view that human sperm each contained a homunculus which directed the formation of the adult organism. His novel *Tristram Shandy* thus opens with the eponymous narrator mourning the "thousand weaknesses both of body and of mind" to which he may be subject as a result of shock to his original homunculus occasioned by an interruption of his conception.

As Sterne admits, the homunculus appeared in a "low and ludicrous light" to some even then, and his place directing development has now been taken DNA. But the thousand weaknesses that can be produced by shocks to the DNA have come to underlie modern developmental biology, and were very much in evidence at *Nature's* conference in Boston this year. With two of this year's Nobel prizewinners present (Christiane Nüsslein-Vollhard, Max-Planck-Institut für Entwicklungsbiologie, Tübingen, and Eric Wieschaus, Princeton), the mood could only be triumphant. But no-one was resting on their laurels.

Occasionally, our understanding of a developmental pathway can be applied to a new system even before mutants are available. Sean Carroll (HHMI, University of Wisconsin-Madison) described how the eyespots on butterfly wings begin as small spots of Distalless, the same protein responsible for proximo-distal axis formation in insect limbs. Only now are mutations affecting the process being found.

But elsewhere, the mutations were inescapable. Nüsslein-Vollhard, indeed, has made a virtue out of necessity by following her prizewinning work on *Drosophila* with an even larger screen for mutations affecting zebrafish development. More than 350 have now been identified, many with intriguing phenotypes, although cloning the genes responsible will not be easy.

In general, though, the focus was tighter. Some studies concentrated on events so early that they are controlled by maternal proteins, so that mutations are evident only in the offspring of an affected mother. In the nematode *Caenorhabditis elegans*, for instance, the search for such mutations is uncovering the molecular apparatus responsible for the asymmetry of the initial cell division which directs subsequent development (James Priess, Fred Hutchinson Cancer Research Center, Seattle).

In *Drosophila*, too, maternal Gurken protein (a TGF α homologue) triggers the signalling by posterior follicle cells which in

turn prompts anterior-posterior axis formation (Daniel St Johnston, Wellcome/CRC Institute, Cambridge, UK). In this case, one result is the formation of a microtubule cytoskeleton, which then directs migration of the oocyte nucleus to one side of the cell where the signalling process is repeated to form the dorso-ventral axis.

Elsewhere, the interest lay in seeing how different organisms employ a known pathway. The Hox gene cluster that regulates anterior-posterior patterning in most (if not all) metazoans is now so well known that it received little comment. But what is its role in *Caenorhabditis*, in which cell lineage dictates virtually all developmental decisions? Although the Hox gene *Mab5* is expressed in a zone towards the rear of the embryo, the cells that express it come from many different lineages (Cynthia Kenyon, UCSF). Can this be a hitherto unheard-of case of positional signalling in nematodes? Remarkably, the answer appears to be no.

Another, equally surprising case is that of dorso-ventral axis formation. Although it is more than 150 years since St Hilaire first suggested that arthropods and vertebrates are inverted with respect to each other, it was only established this year that their dorso-ventral axes are indeed laid down by related but inverted sets of proteins (Eddy De Robertis, UCLA).

The very success of the mutational approach has sometimes obscured how much biochemistry remains unclear. An honourable exception is that of mesoderm induction by fibroblast growth factor (FGF) in *Xenopus*, which appears to reflect induction of the transcription factor Brachyury by the MAP kinase pathway (Jim Smith, NIMR, London). But few events are so well understood. Even the receptors for proteins such as Wingless and Hedgehog are still unknown, although Philip Ingham (ICRF, London) presented a candidate for the latter. Similarly, while many proteins involved in intracellular signalling by Wingless have been identified genetically, others (such as a kinase that interacts with Dishevelled) are emerging only now that biochemical studies have begun (Roel Nusse, HHMI, Stanford).

Sometimes, a molecule's structure reveals surprises. The C-terminal domain of Hedgehog was already known to remove itself by autoprolysis, perhaps tethering the N-terminal domain (which is responsible for signalling) to the cell surface. But the X-ray crystal structure of the N-terminal domain from murine Sonic Hedgehog shows that it too may be a protease, although its substrate remains to be found (Philip Beachy, HHMI,

Johns Hopkins).

In addition to its complex role in notochord and neural tube formation (Beachy), Sonic Hedgehog exerts a concentration-dependent effect in growing mouse limbs, where it is expressed in the zone of polarizing activity (Andrew McMahon, Harvard). But such different roles are often fulfilled by different members of the one protein family. Several different FGFs can induce vertebrate limb formation, for instance, although FGF8 appears to initiate the process in the chick (Gail Martin, UCSF), and different Wnt family members mediate neural crest formation (Marianne Bronner-Fraser, UC Irvine) and dorso-ventral patterning of the limb (McMahon).

Some of the Hox genes, too, are involved in limb formation, although it has proved very difficult to understand how their expression is confined to the relevant parts of the limb (Dennis Duboule, Geneva). Not all the proteins involved are so well known, however, and some are only starting to attract the attention they deserve. The *Fringe* genes, for instance, may play an important and conserved role in boundary-specific signalling (Wieschaus).

But it was the accounts of plant development that drove home how much animals have in common. In a sense, this is not surprising — plants and animals share no multicellular ancestor, making the former the "ultimate outgroup", as Elliot Meyerowitz (California Institute of Technology) put it. But while the differences between the various animal phyla diminished with each talk, and the same players surfaced repeatedly in different roles, here all was new.

Gerd Jürgens (Tübingen) indeed described a mutation affecting axial patterning in *Arabidopsis*, but the gene responsible bore no relation to those fulfilling this function in animals. And while the different parts of flowers (a new type of biological pattern, only 120 million years old) can be homeotically transformed, as Meyerowitz showed, their identities are controlled in a very different way from those of metazoan segments.

By drawing attention to the unity elsewhere, these findings raise the most exhilarating possibility of all: that the basic processes governing the development of all metazoans are the same. As well as its intrinsic intellectual elegance, such a principle would greatly simplify the application of results obtained in one phylum to another. But it also raises some fascinating suggestions. To name just one, might not the Cambrian explosion reflect the first evolution of this bodyplan?

Nicholas Short

On the wings of Pegasus

Gordon Walker

HAS the search for other Jupiters seen success at last? Ever since the eighteenth century when Kant and Laplace put forward their nebular hypothesis for the formation of the Solar System, astronomers have generally assumed that the Sun and planets condensed out of a large, rotating disk of gas and dust, material which had earlier contracted out of the interstellar medium. This hypothesis implies that the orbital motion of planets contains most of the angular momentum of the original

would be a mere 10^{-9} as bright in visible light. Most efforts have been directed towards detecting the periodic reflex motion of the parent star about the common centre of gravity with any unseen planet. One can either make an astrometric measurement of the displacement of the star image against the background of distant stars or use the Doppler effect to measure line-of-sight velocity changes. These two techniques are complementary: the larger the separation between planet

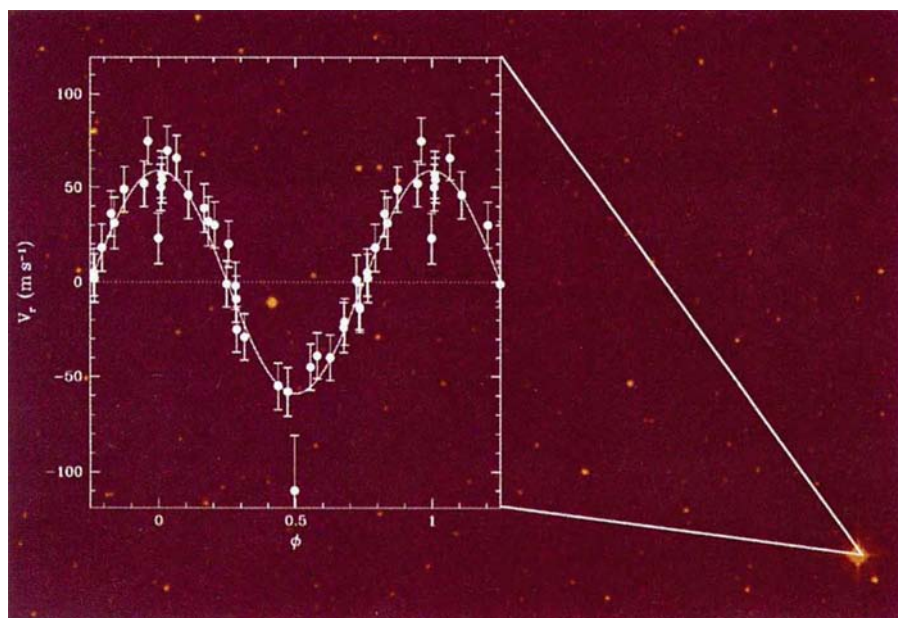
panions with periods of 12 and 26 years. As Barnard's star is one of the closest to the Sun (1.8 parsecs away), this conclusion encouraged the view that systems of giant planets are indeed frequent. Sadly, the detection was not confirmed when later studies were made by Gatewood and others (see ref. 1), and at the moment there are no putative astrometric detections of planets.

Because of Jupiter's influence, the centre of the Sun pivots about a point close to its surface with a period of about 12 years at the modest speed of 13 metres per second. The corresponding Doppler shift at optical wavelengths is only a few ten-thousandths of an ångström. Nonetheless, several teams including the Geneva group have independently developed techniques sensitive enough to detect the reflex signature of a Jupiter-sized planet in radial velocities over many years.

Despite this, a long-term Canadian survey⁴ of 21 single, solar-type stars found no planets to a lower limit of one to three Jupiter masses for periods between 40 days and 15 years. This survey also ruled out companions of 3 to 10 Jupiter masses at much longer periods based on long-term trends in the radial velocities and limits imposed by astrometry. When these results were combined with those of other searches⁵⁻⁷, the number of scrutinized stars without detected Jupiter-mass planets increased to 45. This was indeed surprising because it implied that such planets might be much rarer than expected. (All of these limits assume orbital planes parallel to the line of sight.)

But finally we may have a detection. Moreover, the remarkable feature of the planet discovered by Mayor and Queloz is its orbital period of only 4.23 days, placing it at only 5% of the distance of the Earth from the Sun but with the mass of a gas-giant. None of the surveys cited above had anticipated finding gas-giants in such tight orbits and so were not very sensitive to such short periods. Perhaps there is a population of short-period giant planets waiting to be discovered.

On the other hand, some tantalizing results did emerge from these other studies. A group under David Latham at Harvard found a clear signature⁸ for perturbation by a companion of at least 11 Jupiter masses, possibly a brown dwarf, in a highly eccentric 89-day orbit about the dwarf star HD114762. My own group⁹ found the classical signature of perturbation by a Jupiter-mass planet in a circular 2.5-year orbit about the bright star γ Cephei. In both cases the attribution to a planet was questioned. The high eccentricity for HD114762 and the possibly high inclination of the orbit to the line of sight implied that the companion could well be of stellar mass. For γ Cephei, which has evolved to the giant stage, the changes in chromospheric activity are in phase with



Signature of a planet? Periodic motion of the star 51 Pegasi, recorded by Mayor and Queloz¹.

cloud which would otherwise have prevented the contraction of the Sun to its present size¹. With over half of the stars in the solar neighbourhood being components of double or multiple systems, it seems reasonable to assume that planetary systems are common, if not the norm around single stars.

Until now astronomers have searched in vain for planets around normal stars but, on page 355 of this issue², Michel Mayor and Didier Queloz of the Geneva Observatory announce the discovery of an object half the mass of Jupiter orbiting very close to the star 51 Pegasi. Because 51 Peg is just visible to the naked eye and so closely resembles the Sun, the original announcement in October not only brought 51 Peg to popular attention, it also galvanized the community of professional planet seekers and theorists, particularly as the proposed planet seems to be unexpectedly close to the parent star.

Detection of planets, even gas giants such as Jupiter, is a severe challenge¹. Jupiter is only one-thousandth the mass of the Sun and, viewed from a distant star,

and star, the larger the astrometric displacement and the smaller the change in velocity, and vice versa. But even for a Jupiter-mass companion, such changes are extremely subtle.

In 1992, the radioastronomers Wolsczan and Frail were the first to detect a system of planets³, consisting of two terrestrial-mass planets orbiting the 6.2-ms pulsar PSR1257+12, with periods of months. Subsequently, a third companion of lunar mass was discovered. Pulsars are precise cosmic clocks whose rate can be directly compared with atomic clocks on Earth. Unfortunately for the theorists, this system could not clarify the question of the incidence and nature of planets around normal stars, and the possible role planetary systems play in star formation. Consequently, the incentive to find planets around solar-type stars remains strong.

Using some 60 years of data taken at the Sproul Observatory, van de Kamp published one of the best known astrometric studies of nearby stars (see discussion in ref. 1). He concluded that Barnard's star had two Jupiter-mass com-

Astronomical questions of origin and survival

WITH anything living, or merely astronomical, questions of origin and survival are central. This is no less true for Mayor and Queloz's putative planet, otherwise known as 51 Peg B. At a distance from its primary star of 0.05 astronomical units (1 AU being the distance of the Earth from the Sun), and with surface temperatures that must exceed 1,000 kelvin, is 51 Peg B on its death bed, or merely tanning? Will it evaporate, be ablated by a stellar wind, or persist?

Surprisingly, the case for the survival of a giant planet one hundred times closer to its primary than Jupiter is to the Sun is compelling, even if that planet, like Jupiter, is largely composed of the light elements hydrogen and helium. The obvious possible loss mechanisms are: Roche lobe overflow; classical Jeans evaporation of the high-energy tail of a maxwellian distribution of molecular and atomic hydrogen in the outer envelope; and the production, by stellar ultraviolet radiation, of hot atoms and ions that have sufficient energy to escape. As all objects near Jupiter's mass that are older than 10^9 years have radii that are less than two times Jupiter's (whatever their composition), it is easily shown that 51 Peg B must now be comfortably within its Roche tidal lobe¹. The rate of classical Jeans evaporation is proportional to $\exp\{-\lambda\}(1+\lambda)$ (ref. 2), where $\lambda = \phi/kT$, and ϕ is the depth of the gravitational potential well ($GM_p m_H/R_p$, where p stands for 'planet' and the other symbols have their standard meanings).

For a Jupiter-mass object with Jupiter's composition and radius, λ is 20 eV, and $kT = 0.1$ eV. Clearly, with λ of 50–400, the classical evaporation of

hydrogen and helium objects with masses between 0.5 and 2.0 Jupiter masses and radii less than twice that of Jupiter is insignificant.

The ultraviolet production of hot ions and atoms is a far more promising mechanism for driving a hydrogen wind from a gas giant. Such a process is known to erode (but only slightly) the atmosphere of Jupiter³ and similar processes (with an intermediate photo-lytic step) have affected the atmospheres of Titan, Earth and Venus. Scaling from the case of the Sun and Jupiter, we can estimate that, at 0.05 AU, a Jupiter would lose H_2^+ and H^+ ions and H atoms with a total flux of $3 \times 10^{13} \text{ cm}^{-2} \text{ s}^{-1}$, which translates into $10^{34} \text{ atoms s}^{-1}$ or $\sim 10^{-16}$ solar masses per year. So, over 10^{10} years, perhaps 0.1–1.0% of the planet can be lost, and although a wind of the dissociation and ionization products of molecular hydrogen is expected if 51 Peg B is a gas giant, its survival is assured. (Of course, if 51 Peg B had ever been closer to its primary than it is now, the stripping rate would have been correspondingly larger.) Ablation by a stellar wind at the solar rate may be no more corrosive and a giant terrestrial planet of high-atomic-mass material would be even harder than a gas giant.

So it is the origin, not the survival, of 51 Peg B that puzzles the mind. Could it be a giant planet formed in the usual way at 5 AU, which through early tidal interactions with the protostellar disk lost most of its angular momentum and plunged inwards (D. Lin, personal communication)? Could it be a giant terrestrial planet (with or without an atmosphere)? Could it have nucleated *in situ* around a rock (not rock/ice) core,

but otherwise be a gas giant? Could it have started out as a brown dwarf or star which has since been tidally stripped?

All these possibilities have their merits and problems (some severe), but the last may be the most bizarre. Close binaries are well known and well studied, but mechanisms for eroding a star or a brown dwarf in this context are few: thermal or wind ablation and Roche lobe overflow. By the same arguments as above, the former mechanisms are unlikely, particularly given the much greater gravitational binding energies. But the latter too has its difficulties, as the relevant mass-radius relation forces the companion to start inside its primary if it is to shed more than 95% of its mass by conservative mass transfer and end up as 51 Peg B at 0.05 AU.

Nevertheless, ablation and tidal truncation by the primary may have moulded the early orbital and structural evolution of 51 Peg B. Whatever the provenance of this roasting beast, its discovery should be a lesson to us: as the search for extrasolar planets accelerates, we should expect to be surprised.

Adam Burrows and Jonathan Lunine

Adam Burrows is in the Departments of Physics and Astronomy, and Jonathan Lunine is in the Department of Planetary Sciences, University of Arizona, Tucson, Arizona 85721, USA.

1. Guillot, T., Burrows, A., Hubbard, W. B., Lunine, J. & Saumon, D. *Astrophys. J. Lett.* (submitted).
2. Chamberlain, J. W. & Hunten, D. M. *Theory of Planetary Atmospheres*, 2nd edn (Academic, Orlando, 1987).
3. Acrey, S. K. *Atmospheres and Ionospheres of the Outer Planets and their Satellites* (Springer, Berlin, 1986).

the velocity. This suggests that the velocity variations may be intrinsic to the star itself.

The results for these two stars highlight the problems with the Doppler search technique. First, without additional information, such as an eclipse, one does not know the inclination of an orbit to the line of sight and so one can only quote a minimum mass for the unseen companion. Second, unlike pulsars, normal stars are not clocks or precise wavelength standards. They ring mechanically like huge bells in a broad range of frequencies, and the evolved stars can be unstable over periods ranging from hours to years. Further, the large upwelling flows of the type that cause the granulation structure on the Sun, together with the presence of magnetic fields, star spots and rotation, can generate cyclical variations in radial velocity.

Is the new detection solid? Given the

short period and comparatively large velocity amplitude that Mayor and Queloz find for 51 Peg, there will undoubtedly be intensive spectroscopic studies to see whether intrinsic velocity variability can be ruled out. In favour of the planet interpretation, the velocity curve (see figure) is remarkably sinusoidal and stable, a point confirmed by groups in Colorado and at the Lick Observatory, California.

There is no question that the short period and large mass of the proposed planet are a challenge to theorists. As discussed in the accompanying box, there is a case for the survival of a gas giant so close to a solar-type star. But how did it get there in the first place? Would an infrared signature from such a hot planetary atmosphere be detectable in the spectrum of 51 Peg itself?

Still more interesting, perhaps, is the second-order trend in the velocities which

could be due to a second gas-giant in an orbit of a few years. This could mean that 51 Peg actually has a system of planets. Now that would be exciting indeed! □

Gordon Walker is in the Department of Geophysics and Astronomy, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada.

1. Black, D. A. *Rev. Astr. Astrophys.* **33**, 359–380 (1995).
2. Mayor, M. & Queloz, D. *Nature* **378**, 355–359 (1995).
3. Wołszczan, A. & Frail, D. A. *Nature* **355**, 145–147 (1992).
4. Walker, G. A. H. et al. *Icarus* **116**, 359–375 (1995).
5. Marcy, G. W. & Butler, R. P. *Publ. astr. Soc. Pacif.* **104**, 270–277 (1992).
6. Cochran, W. D. & Hatzes, A. P. in *Planets Around Pulsars* (eds Phillips, J. A., Thorsett, S. E. & Kulkarni, S. R.) 267–274 (1993).
7. McMillan, R. S., Moore, T. L., Perry, M. L. & Smith, P. H. *Astrophys. Space Sci.* **212**, 271–280 (1994).
8. Latham, D. W. et al. *Nature* **339**, 38–40 (1989).
9. Walker, G. A. H. et al. *Astrophys. J.* **396**, L91–L94 (1992).

neu tack on neuregulin

Mark A. Marchionni

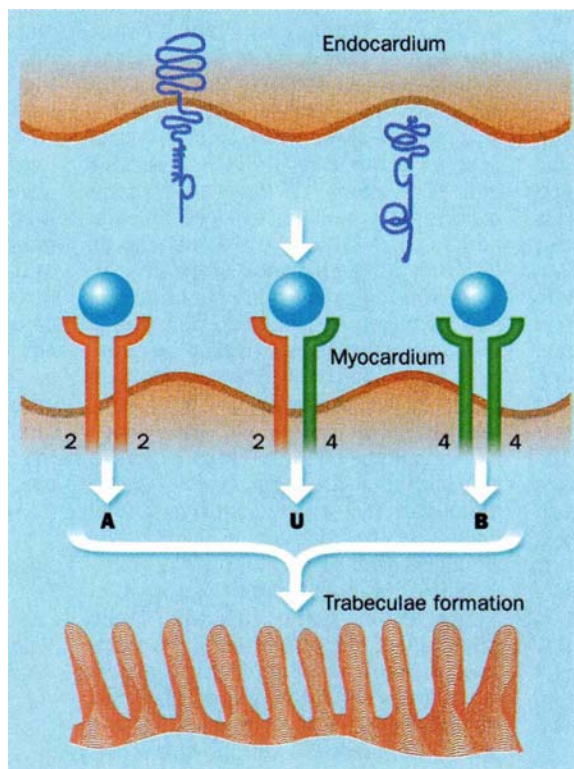
SIGNALLING by peptide growth factors through receptor tyrosine kinases is central to many of the cell-cell interactions that regulate embryonic development, tissue maintenance and repair. Aberrant expression can lead to tumour formation; inadequate expression can disrupt embryonic development or result in degenerative disease in the adult.

One such system is that composed of the neuregulin growth factors and their receptors, erbB2 (also called *neu*), erbB3 and erbB4. As reported in three papers in this issue¹⁻³, three of the genes concerned have been disrupted in mice — those encoding neuregulin¹ and erbB2 (ref. 2) and erbB4 (ref. 3). The results obtained are both gratifyingly congruent and interestingly divergent. All three genes are essential for development, the mutant embryos dying *in utero* at day 10.5 and displaying similar heart malformations. The embryos also have defects in the nervous system that overlap, but are not identical.

Over the past few years, thanks to the identification of some key molecules, a clearer picture of neuregulin-erbB interactions has emerged^{4,5}. Neuregulins are a large group of secreted and membrane-attached growth factors, expressed as alternatively spliced isoforms from a single gene. Members of the group (for example *neu* differentiation factor, heregulins, glial growth factors and acetylcholine-receptor-inducing activity) are produced by neurons and mesenchymal cells, and elicit various trophic responses (proliferation, survival, differentiation) in a wide range of cell types. Neuregulin actions are mediated through one or more tyrosine kinases of the epidermal-growth-factor (EGF) receptor family. It was once thought that neuregulins acted solely through erbB2, but the situation has become less straightforward. From the evidence available now⁵, it seems that erbB2 binds neuregulins with high affinity only with the aid of a co-receptor, and that erbB3 binds effectively alone, but is unable to transduce a signal. Hence, erbB2 and erbB3 are thought to collaborate as a heterodimer; erbB4, on the other hand, is capable of both binding and signalling and thus may function independently.

The new results show that, in the devel-

oping rodent heart, neuregulin expression is confined to the endocardium¹. Myocardial cells express both erbB2 (ref. 2) and erbB4 (refs 1, 3). Expression of erbB3 is localized to the endocardial cushion¹. Mice lacking neuregulin, erbB2 or erbB4 appear to have identical defects in heart development: the trabeculae (finger-like



Neuregulins are expressed in the endocardial cells of the heart¹ either as secreted growth factors (right) or as membrane-attached forms (left), which are subsequently released (clipped) from the cell surface; all forms of neuregulin contain an EGF-like fold and most of them an immunoglobulin-like domain. This paracrine neuregulin signal (shaded blue sphere) is received at nearby (possible apposing) myocardial cells, which bear the receptors erbB2 (2) and erbB4 (4). On binding neuregulin, the receptors dimerize, then activate intrinsic kinase activities to initiate the cascade of intracellular signal transduction. Because erbB2 and erbB4 are coexpressed, and both genes are required for the formation of the fingerlike trabeculae essential to cardiac development, this developmental process must be driven by the union of signals (AUB) resulting from independent activation of the two homodimeric receptors or from their putative formation into a heterodimer. The specific signal-transduction pathways leading to myocardial trabeculation need to be worked out in more detail.

extensions of the ventricular myocardium) are absent¹⁻³ (see figure). Elsewhere the myocardium looks normal, including the atrial and ventricular walls, where erbB2 and erbB4 expression is readily detected. The phenotype of the developing heart in erbB3 null mutants is anxiously awaited.

The neural defects analysed at embryonic day 10.5 in these three mutant mice are restricted to the developing rhomben-

cephalon, a brain structure that connects the cervical spinal cord to the midbrain and develops into the brainstem. In the rhombencephalon, development follows a highly segmented pattern — the neural tube is subdivided transiently into rhombomeres (r1-r8) that give rise to both motor neuron pools and migrating neural crest cells⁶.

Analysis of embryos from days 9–10.5 of development shows that the patterns of expression and the neural phenotypes of neuregulin and erbB2 gene disruptions are consistent with a ligand–receptor relationship operating by paracrine signalling (that is, signalling between nearby cells)^{1,2}. In general, neurons provide the neuregulin signal(s) and glia or other apposing target cells respond through erbB2 activation. Both mutants lack the neural-crest-derived portions of cranial sensory ganglia. Neuregulin mutants are also devoid of Schwann cell precursors along peripheral nerve roots from the spinal cord, which may be due to neuregulin effects on proliferation, survival or cell-lineage commitments^{4,7,8}. Further, erbB2 (and presumably neuregulin) null embryos lack both sensory and motor innervation connecting cranial ganglia to corresponding rhombomeres. Placode-derived sensory neurons of the trigeminal ganglia, though still present, also fail to innervate the hindbrain. In cell culture studies, neuregulins expressed in neuronal precursors have been shown to upregulate the level of neurotrophin-3 activity secreted from non-neuronal, neural-crest-derived cells of embryonic sympathetic ganglia (J. Verdi, personal communication).

Taken together, these observations suggest that disruption of a reciprocal cell–cell trophic-support circuit may account for much of the neural phenotype in neuregulin- and erbB2-deficient mice. An autocrine mechanism, in which the signal affects the cell generating it, cannot be ruled out, however. Interestingly, survival and neurite extension of embryonic retinal ganglion cells (which express both neuregulins and erbB receptors)

is promoted by neuregulin⁹. A more refined map of neuregulin and erbB expression patterns in the rhombencephalon will help clarify how the signals are transmitted.

The evidence³ of erbB4 expression and function in the rhombencephalon is strikingly different to that from neuregulin and erbB2. Mutation of erbB4 does not result in the loss of cells and axons of the cranial

sensory ganglia, as observed in the other mutants. Rather, the organization, spacing and pattern of innervation of these ganglia to and from the central nervous system is disrupted. Mistargeting is prevalent, as axons fail to recognize normal chemorepulsive boundaries and to adopt normal trajectories. This finding represents the first piece of genetic evidence that supports the idea of chemorepulsion of axon growth in vertebrates and reveals a new and unexpected role for erbB4. Just how erbB4 does this is another mystery to be solved — thus begins the screening of candidate chemotropic molecules¹⁰ whose expression might be regulated by erbB4. Because this phenotype is not seen in neuregulin-deficient mice, Gassmann *et al.*³ also suggest that erbB4 responds to a different, structurally related ligand at this time and place in development.

The convergence of the three gene knockouts now brings into focus some key features of the biology of neuregulins and erbB receptors. These initial studies implicate another ligand, and there may be a suitable candidate on the horizon¹¹; erbB4 seems to operate less independently as a neuregulin receptor than previously presumed; and the erbB2–neuregulin relationship is revitalized. Indeed, what constitutes a neuregulin receptor takes on greater meaning in the context of these genetic data than might be concluded from other approaches.

Future use of the gene-knockout approach to examine how neuregulins and erbB receptors regulate development and function will depend on extending embryo viability by transgenic rescue or by targeting specific tissues. Knockouts of individual exons of the neuregulin gene may have subtler phenotypes, which can provide clues about specific functions of the many isoforms. The future genetic experiments will be able to draw on what is already known about sites of expression mapped in more mature embryos and in adults, as well as knowledge of cellular activities gleaned from culture studies.

Oncologists have recognized that overexpression of HER-2, the human homologue of erbB2, correlates with poor prognosis in breast cancer (see ref. 5), and several HER-2 antagonists are currently undergoing clinical trials. Now we

are aware that neuregulin–erbB signalling is essential for development of the heart and nervous system. Sustained expression of these genes after birth and into the adult suggests that they perform functions vital for tissue maintenance and repair. Some degenerative disorders, such as cardiomyopathies, neuropathies and demyelinating diseases, might be alleviated

by augmenting neuregulin–erbB signalling, meaning that the neuregulins have potential as drugs for treatment of neuromuscular, neurological and sensory disorders. □

Mark A. Marchionni is at Cambridge Neuro-Science, Inc., One Kendall Square, Building 700, Cambridge, Massachusetts 02139, USA.

SOLVATION

Hopes for Hofmeister

V. Adrian Parsegian

It is not usual for researchers using as advanced a technology as the ISIS pulsed-neutron source to cite a nineteenth-century reference as part of the motivation for their work, yet that is what Leberman and Soper¹ do on page 364 of this issue. After more than a century, the mysterious trends of the Hofmeister series² continue to challenge us. Originally seen through the specificity of different ions in precipitation, or 'salting-out,' of proteins, these trends permeate a staggering number of phenomena. A thought-provoking 1984 compilation by Collins and Washabaugh³ mentions well over 30 properties of salt solutions that show systematic differences with ionic, particularly anionic, type: the stability, precipitation and solubility of proteins; transport, spectral and thermodynamic properties of water and salt; solution surface potentials and surface tensions, to name but a few.

What decides whether an ion stabilizes or destabilizes a molecular surface? What is the magic or secret of a typical sequence (say, sulphate-phosphate-fluoride-chloride-bromide-iodide-perchlorate-thiocyanate) such that so many properties vary as a monotonic function of this list? The best guess is that one should focus on the solvation effects of these ions. More specifically, Hofmeister series are thought to reflect different ordering powers of ions, usually anions, on the surrounding water molecules. The ionic sequence goes from stabilizing 'cosmotropes' to (relatively) disruptive 'chaotropes'. The challenge is to go from establishing macroscopic, phenomenological consistencies to microscopic examination of what the different ions do to water.

In this issue Leberman and Soper¹ actually look at the water itself in concentrated salt solutions. They compare neutron diffraction patterns from four different salts dissolved in D₂O and H₂O to extract the distribution of distances $\Delta g_{\text{HII}}(\mathbf{r})$ between water protons in the presence of salts. (The use of different hydrogen isotopes lets them identify contributions from protons only.) The point of focus here is the number of proton pairs at the 2.4-ångström H–H separation

occurring across an intervening oxygen. The H...OH peak in the distribution function is taken as the sign of a hydrogen bond. This 'reporter' peak is changed by added salts; so are the features of the distribution functions many ångströms away. It is difficult to see data such as these and think that solvation of ions or polar molecules is effectively restricted to a first coordination shell, or hydrated radius, as is still sometimes assumed.

What is remarkable in the work by Leberman and Soper is that after subtracting all the blanks and taking all the differences-of-differences in scattering patterns from concentrated solutions of different salts and isotopes, it is possible to extract accurate proton–proton correlation functions with enough features to speak of the changed weight of diffraction peaks. Earlier neutron diffraction reinforced the general idea that water 'structure' results from many-body cooperative or collective interactions. The present data show the extent of correlation by focusing precisely on H-bonding.

From visual similarities in the distribution of H–H separations, the authors suggest that salts have a disordering effect similar to that of pressure (or density). But heating also breaks the H-bond network. Why pick pressure? The pure-water measurements here are done as a function of pressure, not temperature. In some sense pressure is the natural variable given the crushingly high electrostrictive forces emanating from concentrated ions. It is a reasonable connection also from the perspective of simulations suggesting that lowered density makes it possible to form more H-bonds⁴.

The correlation that the authors aim to draw between suppression by different ions of the 2.4-Å H...OH peak and what these different ions do to proteins is far more ambitious. Sylllogistically:

- The Hofmeister series is a sequence in water-ordering power of ions.
- Neutron diffraction shows that ions modify the H...OH form to different degrees.
- Neutron diffraction gives a microscopic measure of the Hofmeister effect.

1. Meyer, D. & Birchmeier, C. *Nature* **378**, 386–390 (1995).

2. Lee, K.-F. *et al.* *Nature* **378**, 394–398 (1995).

3. Gassmann, M. *et al.* *Nature* **378**, 390–394 (1995).

4. Lemke, G. *Nature* **362**, 292–293 (1993).

5. Carraway, K. L. & Burden, S. J. *Curr. Opin. Neurobiol.* **5**, 1–7 (1995).

6. Lumsden, A., Spradon, N. & Graham, A. *Development* **113**, 1281–1291 (1991).

7. Dong, Z. *et al.* *Neuron* **15**, 585–596 (1995).

8. Shah, N. M., Marchionni, M. A., Isaacs, I., Stroobant, P. & Anderson, D. J. *Cell* **77**, 349–360 (1994).

9. Bermingham-McDonogh, O., McCabe, K. & Reh, T. *Soc. Neurosci. Abstr.* **21**, 547 (1995).

10. Keynes, R. & Cook, G. M. *W. Cell* **83**, 161–169 (1995).

11. Chang, H. & Gilbert, W. *Soc. Neurosci. Abstr.* **20**, 1694 (1994).

But syllogisms have their limits. Changes in proton distribution are not in themselves equivalent to changes in protein stability. Nevertheless there are some new pieces here that have been missing from an old puzzle. We can think about the action of salts on large molecules and surfaces; we can heed the analogy to pressure made by Leberman and Soper.

A common theme in thinking about the many effects of ions is the competition for water between ionic solutes and the larger molecular or macroscopic surfaces. But there is much more to it than that. The same thirsty ions that dehydrate or precipitate proteins in one situation can also have an affinity for proteins in other cases. The same salt that precipitates a protein at one concentration can 'salt it in' at another⁵. We really don't understand enough about ions near surfaces or near macromolecules to explain how they compete near these surfaces. It has become clear that interactions between solute and protein are a specific and delicate balance of adsorption or binding and exclusion. The decision of an ion to act on a surface has a direct-binding part and a solvation-sphere part⁶. There is as much reason to

seek a Hofmeister series for ion exclusion as for ion binding.

After all, proteins, as much as ions, are likely to do their own rearranging of water. Supple, flexible, adjustable networks of H-bonds fill the space between collagen fibrils brought to the densities they have *in vivo*⁷. This is the kind of rearrangement that one expects from 'hydration forces' measured between such fibrils, an interaction that shows strong evidence of salt exclusion near the protein as well as a force dominated by macromolecular solvation⁸. In another microscopic glimpse of ion-water association competing with ion-peptide association, electrical current fluctuations have shown that anions occupy sites on roflamycin channels for different characteristic times in a preference sequence that follows a Hofmeister series⁹.

Just how does one think about all this? There are the attractions and repulsions between ions and surfaces, ions and water, water and surfaces, in every combination. Is ion binding driven by its activity, as one would think of an adsorption isotherm? It can also be driven by the release of water liberated from charged solute and binding site. Air-water surface potentials are

sometimes a linear function of concentration rather than logarithmic in concentration and are not therefore straight functions of ion activities. Linearity here suggests that it is the work of releasing ion-structured or surface-structured water that is the determining factor. It is as though we should think of ion approach to a surface as an osmotic event rather than a direct binding or repulsion.

Another possible implication of the report by Leberman and Soper is the analogy to be pursued between the action of hydrostatic pressure reorganizing water and the reorganization effected by inserting ions into water. Again, there is a very old literature connecting hydrostatic pressures, solution volumes, ion association and macromolecular stability. The immediate picture drawn from Le Chatelier's principle is that applied pressure favours the states of higher density. Charges pull in the water dipoles; under high pressure, ions are less likely to surrender the water compacted around them. The next step is to think more generally about pressure destabilization of bonding patterns in all interactions. Weber¹⁰ has recently suggested that pressure can act differentially to destabilize bonds involving water with itself, water with the protein or macromolecular surface, and surfaces with each other. Each is a disturbance from a most-favoured low-pressure structure. Their sum is more than the action of pressure on water alone, even as ion effects are more than the action of ions on water alone.

At least in the situations where we encounter it, water is no ordinary liquid, nor is its encounter with ions or macromolecules a simple event. The hope raised by neutron diffraction measurements is that there can be a merger of macroscopic and microscopic strategies to see correlations in a different perspective — and maybe make some good references for a paper in *Nature* (in 2088 or 2095). □

V. Adrian Parsegian is in the Division of Computer Research and Technology, and the National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA.

NEUROBIOLOGY

Nitric oxide and bad behaviour



An onlooker watches as two male mice with disrupted neuronal nitric oxide synthase (nNOS⁻) genes become locked in combat. Solomon Snyder and co-workers had established a breeding colony of nNOS⁻ mice, and the mysterious and apparently violent deaths of males housed together suggested to them that lack of nNOS, and hence of neuronal nitric oxide, might be connected to a tendency to exceptional aggressiveness. The mice seemed otherwise behaviourally normal.

Snyder and colleagues took a closer look at these mice, and describe their observations on page 383 of this issue. When encountering other male mice, nNOS⁻ males are less likely to be submissive and more likely to initiate and persist in violent attacks than wild-type males. When caged with female mice, they continue to mount non-oestrous females even when the females show no interest in sexual encounters. Female nNOS⁻ mice, however, behave normally.

Nitric oxide signalling is important in blood vessels and macrophages, and in neurotransmission. But until the advent of gene knockout technology it was not possible to investigate what behavioural functions the neuronal form might have, because NOS inhibitors perturb various aspects of mouse physiology.

The knockout mice have normal neuronal plasticity and normal levels of testosterone. So why are they so aggressive? One possibility is that their aberrant behaviour is due to disruption of nNOS-containing neurons in the limbic system, an area of the brain involved in the generation of emotional behaviours. Other explanations remain tenable, however, and only if that connection is confirmed will it be clear whether studies on humans will be of interest. Harriet Coles

1. Leberman, R. & Soper, A. K. *Nature* **378**, 364–366 (1995).
2. Hofmeister, F. *Arch. Exp. Path. Pharmacol.* **24**, 247–260 (1888).
3. Collins, K. D. & Washabaugh, M. W. *Q. Rev. Biophys.* **18**, 323–422 (1985).
4. Sciortino, F., Geiger, A. & Stanley, H. E. *J. chem. Phys.* **96**, 3857–3865 (1992).
5. Green, A. A. & Hughes, W. L. *Meth. Enzym.* **1**, 67–90 (1952).
6. Timasheff, S. N. *A. Rev. Biophys. biomol. Struct.* **22**, 67–69 (1993).
7. Bella, J., Brodsky, B. & Bereman, H. M. *Structure* **3**, 893–906 (1995).
8. Leikin, S., Rau, D. C. & Parsegian, V. A. *Nature struct. Biol.* **2**, 205–210 (1995).
9. Grigorjev, P. A. & Bezrukov, S. M. *Biophys. J.* **67**, 2265–2271 (1995).
10. Weber, G. *J. phys. Chem.* **97**, 7108–7115 (1993).

The thrombin paradox

John H. Griffin

THE thrombin paradox is like the so-called French red wine paradox¹: too much thrombin or red wine is harmful to blood vessels, but a little bit of either is much better than none at all. On page 413 of this issue², Gibbs *et al.* deconstruct the thrombin paradox by showing that the mutation of glutamic acid to alanine at position 229 of the protein's sequence, creating Ala-229 thrombin, dramatically alters its specificity, both *in vitro* and *in vivo*. Thrombin can both promote and prevent blood clotting, and the result of the mutation is that thrombin's procoagulant activities almost disappear whereas its anticoagulant activity is conserved.

Normal blood clotting is essential to minimize bleeding and achieve haemostasis; excessive clotting contributes to thrombosis, which can have severe consequences (stroke or heart attack, for instance). Maintaining normal blood flow requires keeping a systemic imbalance between the anticoagulant and procoagulant pathways, with the former prevailing. Thrombin is central to both processes (see Fig. 1, and ref. 3 for review). Low levels⁴ generate markedly increased levels of the endogenous circulating anticoagulant serine protease, activated protein C (aPC)⁵. In contrast, high levels of thrombin cause blood to clot by generating procoagulant cofactors Va and VIIIa, by converting fibrinogen to fibrin and by activating platelets that then unveil concealed cell-surface receptors and provide various prothrombotic agents. Accordingly, a J-curve describes the relationship between the thrombotic potential of blood and thrombin concentration (Fig. 2). (Because the affinity of wild-type Glu-229 thrombin for thrombomodulin is very high, and there is an abundance of thrombomodulin in the microcirculation, low levels of thrombin have an antithrombotic effect. At higher levels, however, free thrombin exerts its various prothrombotic activities. As Fig. 2 shows, the thrombotic-potential profile of the Ala-229 mutant is markedly different.)

A natural 'protein C knockout experiment' tells us a great deal about the physiological importance of aPC. New-born infants totally deficient in protein C suffer from extensive thrombosis with inflammation in the microvasculature. The condition is lethal unless treated

aggressively, and most infants suffer irreversible loss of vision and brain damage. Activated protein C circulates at 40 picomolar concentration in human blood and has a remarkably long half-life of 0.3 hours⁶, suggesting that it is part of a systemic anticoagulant and anti-inflammatory surveillance system.

Animal experiments have helped to round out the picture. Infusions of aPC

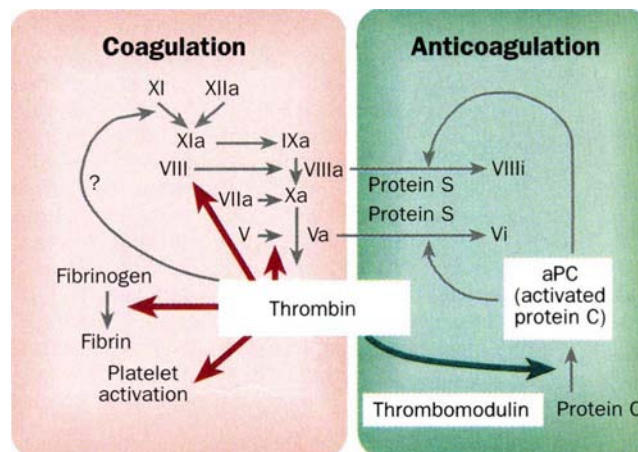


FIG 1. Blood coagulation and anticoagulation (protein C) pathways. Thrombin can be either procoagulant (red arrows) or anticoagulant (green arrows), depending on the substrate. In the first instance, free thrombin activates platelets, clots fibrinogen, and converts inactive forms of coagulation factors V, VIII, and XI to their active (a) forms. But when bound to the cell-surface transmembrane protein, thrombomodulin, thrombin's procoagulant properties are neutralized and its ability to activate protein C is tremendously enhanced. Activated protein C (aPC) is a potent anticoagulant that inactivates factors Va and VIIIa with assistance from the cofactor, protein S, yielding inactive (i) factors Vi and VIIIi.

prevent the lethal outcome of administering *Escherichia coli* bacteria to baboons⁷, and the protein has significant anti-inflammatory activities^{3,7,8}. When patients with acute myocardial infarction receive thrombolytic therapy, circulating aPC is increased about 10-fold. Restricting blood flow in a coronary artery in pigs induces aPC formation and infusions of aPC minimize any consequent heart damage. Infusions of aPC are antithrombotic in baboons bearing prothrombotic arterial shunts⁵ and in dogs when coadministered with urokinase for blockage of the coronary artery. All in all, then, aPC is a normal and potent component of the blood's host-defence system.

Gibbs and colleagues' finding that Ala-229 thrombin elevates levels of aPC *in vivo*, without apparent increase in prothrombotic potential, is a noteworthy achievement and provides a favourable modification of the J-curve that describes thrombotic potential (Fig. 2). This detailed characterization of Ala-229 thrombin activity follows a seminal report⁹

that dissected thrombin's procoagulant and anticoagulant activities by site-directed mutagenesis and inspired an alanine-scanning study of 62 thrombin mutants¹⁰. Whether the new work² should be considered an affirmation of industrial-strength alanine scanning is left to the reader's judgement.

In any case, Gibbs *et al.* provide an appealing *post factum* rationale for the altered selectivity of Ala-229 thrombin. It seems that fibrinopeptide A, one of the two pairs of peptides in fibrinogen cleaved by thrombin (see Fig. 1), is in a conformation folded at the P5 residue, glycine, which interacts with glutamic acid at thrombin position 229 when bound to the thrombin's active-site region. But the P5 residue on protein C is glutamine not glycine, meaning that thrombin interacts with protein C in extended, not folded, form. Gibbs *et al.* suggest that whereas Glu-229 in thrombin is essential for recognizing fibrinogen, it can be dispensed with for recognition of protein C by the thrombin-thrombomodulin complex. Hence the differential effect caused by the mutation.

Given that thrombosis causes so many cardiovascular and other diseases, the need for effective antithrombotic drugs is self-evident. Ideally, these agents should provide their activity without increased risk of bleeding, but no such drug is available; there is a risk/benefit ratio to be considered because of drug-induced haemorrhage that may cause stroke or death. Some low-molecular-weight

thrombin inhibitors, including hirudin and argatroban, are currently undergoing clinical trials. If these inhibitors neutralize the abilities of thrombin to be both procoagulant and anticoagulant, they might not prove clinically useful.

The clinical potential of Ala-229 thrombin remains to be explored, but what is the future of engineered proteins such as this? Ala-229 thrombin could be further engineered to increase selectivity for anticoagulant rather than procoagulant substrates. A promising thrombin or

1. Renaud, F. & de Logeril, M. *Lancet* **339**, 1523-1526 (1992).
2. Gibbs, C. S. *et al.* *Nature* **378**, 413-416 (1995).
3. Esmon, C. T. *Arterioscl. Thromb.* **12**, 135-145 (1992).
4. Hanson, S. R. *et al.* *J. clin. Invest.* **92**, 2003-2012 (1993).
5. Gruber, A. *et al.* *Circulation* **84**, 2454-2462 (1991).
6. Heeb, M. J. *et al.* *J. biol. Chem.* **266**, 17606-17612 (1991).
7. Taylor, F. B. Jr *et al.* *J. clin. Invest.* **79**, 918-925 (1987).
8. Grey, S. T. *et al.* *J. Immun.* **153**, 3664-3672 (1994).
9. Wu, Q. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **88**, 6675-6679 (1991).
10. Tsiang, T. *et al.* *J. biol. Chem.* **270**, 16854-16863 (1995).

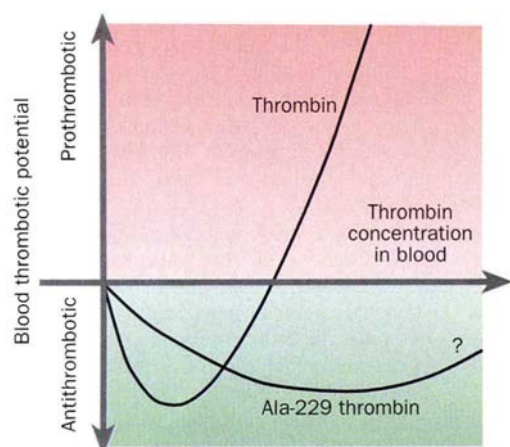


FIG. 2 Influence of the concentration of wild-type and Ala-229 thrombin on the thrombotic potential of blood. Ala-229 thrombin greatly favours thrombomodulin-dependent activation of protein C, and hence antithrombotic potential. The effects of very high concentrations of Ala-229 thrombin are unknown. The scale for each axis is in arbitrary units.

aPC mutant could be altered to prolong *in vivo* half-life by active-site acylation, which could be accomplished using reversible acylating agents that slowly undergo deacylation to release active enzyme over an extended time. The half-life of engineered thrombin might also be prolonged by its encapsulation, for instance in liposomes.

An engineered protein C or aPC with increased resistance to plasma protease inhibitors, but without loss of antithrombotic activity, would have a half-life of

more than 0.3 hours and increased potency. Protein S, a cofactor in the protein C pathway, might be also be engineered for superior antithrombotic activity. Testing and implementation of such engineered agents for clinical usage is limited by the high cost of expressing the vitamin-K-dependent proteins involved in blood coagulation, because of the requirement for mammalian cells to ensure correct post-translational modifications. If genes that accomplish these modifications (for instance those encoding γ -carboxylase) were cotransfected into less expensive expression systems such as bacterial, yeast or insect cells, then costs of recombinant antithrombotic proteins could become more favourable.

The work by Gibbs *et al.*² shows that it is possible to deconstruct and then reconstruct the throm-

bin paradox with a significant net benefit of increased antithrombotic potential. This should stimulate research to discover other engineered proteins or small agents that enhance the activities of the protein C pathway. □

John H. Griffin is in the Department of Molecular and Experimental Medicine, and the Department of Vascular Biology, The Scripps Research Institute, La Jolla, California 92037, USA.

METHANE

An open or shut case?

Everett L. Shock

THE most dramatic evidence for large-scale movement of crustal fluids is the accumulation of ore deposits and petroleum reservoirs. It is a curious twist of fate that these systems are studied separately in the Earth sciences, despite their proximity in many sedimentary basins. Indeed, it is rare that the same researchers have studied both types of systems, or given more than passing reference to one or the other. This separation of economic geology (focused almost exclusively on ore deposits) from petroleum geology has led to conflicting models for the movement of fluids and the transformation of geological materials. The dichotomy even affects the verbs used to explain the processes: metals are 'transported', but petroleum 'migrates'. Now comes evidence from Price and Schoell, writing on page 368 of this issue¹, that the processes in source rocks are not the same as those inferred from studies of petroleum and natural gas, raising the spectre that an ore-deposit view of petro-

leum might be an essential part of the puzzle.

Economic geologists generally accept the notion that the main controls on ore composition are the path of fluid migration and the conditions at the site of deposition. Although equally important, the source of the metals that become concentrated in an ore deposit is rarely known with precision. On the other hand, those who study petroleum deposits (including reservoirs of natural gas) generally agree that composition in the reservoir stems from the composition of the source material and the conditions prevailing in the source rock when the petroleum was generated, despite the fact that the source can be at a considerable distance from the reservoir and often is not precisely located.

This is not the case in the Bakken formation of North Dakota. Here the source rock shales are also the reservoir, and petroleum has remained in place since it was formed. Price and Schoell¹ have

analysed natural gas produced with the petroleum, and their results indicate that the distribution of gases is strikingly different from gases in 'normal' gas reservoirs. Specifically, the data show that the methane content of these gases is conspicuously lower than that of most other natural gas deposits.

Rather than thinking of the Bakken as some sort of anomaly, Price and Schoell suggest that the gas concentrations are truly representative of the original composition of most natural gas, and that the ratios of light hydrocarbons in reservoirs are influenced by the pathways and processes through which the natural gas was transported and accumulated in the crust. It is these processes, then, that are proposed to cause the predominance of methane. Price and Schoell cite as corroborating evidence the results of hydrous pyrolysis experiments using Bakken shales that show similar distributions of light hydrocarbons (especially relatively low methane). If we follow this argument a few steps further, it is logical to propose that it may be extremely difficult to infer the processes occurring in source rocks from petroleum or gas compositions, which are likely to have been altered dramatically by the transport and accumulation processes. If so, large parts of the foundation of petroleum geology may be constructed on shaky ground.

But before we jump in with both feet, we should test the water to see if it is a mirage. Frank Mango and co-workers find that they can generate gas mixtures very much like most natural gas samples if they use transition metal catalysts², or metal-rich source rocks³, in pyrolysis experiments. In fact, they argue that the failure of other pyrolysis experiments to match most natural gas compositions can be blamed on the missing catalysts in the experiments. Does this mean that the Bakken is anomalous, and the reason for its peculiarities lies in the apparent inhibition of transition-metal catalysts?

What none of these investigators states explicitly is that the progress of reactions can be driven by the degree to which a system is open: that is, the extent to which reaction products can leave the system and new reactants enter. Recognition of this fact in the study of ore deposits⁴ provides the conceptual framework for relating theoretical mass-transfer calculations to actual geological samples and processes.

Recent advances^{5,6} in incorporating fluid-flow calculations into geochemical reaction path models are helping to quantify the consequences of fluid/rock reactions in a wide variety of geological settings. The task of including organic compounds in these calculations is being aided by growing experimental⁷ and theoretical⁸ work. It will soon be routinely possible to treat inorganic and organic

transformations in sedimentary basins in the same geochemical models. Initial efforts of this type⁹ have revealed connections between fluid/rock and fluid/organic reactions, and led to yet another hypothesis about the origin of methane in natural gas. In this model, the irreversible step in the transformation of organic compounds is found to be the generation of methane in what is, by necessity, an open system owing to the inevitable fracturing of sedimentary rocks. The leaky system drives additional reversible hydrolytic disproportionation reactions among other hydrocarbons in petroleum. All of these reactions involve water explicitly and imply that there is potential for continuous transformation of hydrocarbons from the source rock to the reservoir.

Meanwhile, microorganisms that produce methane (hyperthermophilic methanogens) have been cultured from various high temperature environments, with growth demonstrated in the laboratory to at least 110 °C (ref. 10). Little is known about the mechanisms of methanogenic reactions at high temperature, or the isotopic fractionation that would accompany these processes. It is nevertheless highly probable that living organisms could be responsible for a larger portion of the methane in natural gas than is generally appreciated. Perhaps the Bakken system, which may be closed to the introduction of new fluids and therefore nutrients, represents an abiotic system. That would at least be consistent with the hydrous pyrolysis experiments. Other 'normal' gas accumulations may contain the kinetic residue of numerous catalytic processes involving mineral surfaces, transition metals and metabolic systems, and may be more renewable than we generally believe. In any event, it is evident that we are still largely ignorant about the processes that provide the fuel on which our society depends. □

Everett L. Shock is in the Department of Earth and Planetary Sciences and McDonnell Center for the Space Sciences, Washington University, St Louis, Missouri 63130, USA.

Thanks for the memory

Nasser Peyghambarian

OPTICAL data storage, where photons play a role in the storage of information, has the ability to reach extremely high densities. Magneto-optical disks are commercially available and have reached densities of around 0.65 Gbits per square inch (0.1 Gbit cm⁻²). In a recent paper, Lin, Wang and Mossberg¹ have demonstrated an areal density of 8 Gbits per square inch for an optical storage system based on a frequency-selective process. Eight billion bits is equivalent to half a million double-spaced typewritten pages — a stack of papers some 50 metres high. The storage process used by this group is a form of spatial-spectral memory in which a single memory cell can store a large number of optical bits that have slightly

different colour or optical frequency.

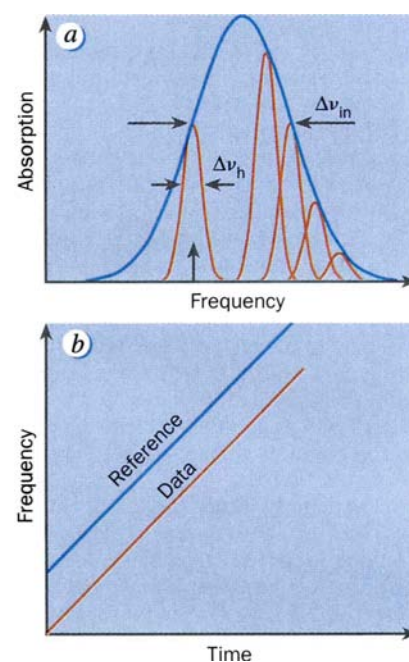
The storage mechanism is described in detail in the accompanying box. Data are encoded as a series of ones and zeros, which then modulate an optical beam (referred to as the 'data' beam) as a stream of laser pulses. As shown in the figure overleaf, an additional reference beam is directed towards the storage material at a small angle to the data beam. The illuminated spot on the storage material constitutes a memory cell that contains all of the data. The information stored may be retrieved later by sending a read beam, which is similar to and in the same direction as the reference beam. The retrieved data are obtained as a stream of pulses similar to the input data in a signal

Spatial-spectral selective optical memory

A MATERIAL with individual atomic or molecular absorbers has an inhomogeneously broadened absorption line that can in principle be used for frequency-selective optical data storage. Consider a material with an absorption profile as shown in part a of the figure. The total spectral width of the absorption profile, $\Delta\nu_{in}$, represents the inhomogeneous linewidth of the sample, while the spectral width of the individual transitions inside the profile, $\Delta\nu_h$, represents the homogeneous linewidth. A laser beam with frequency matching one of the subtransitions, as at the frequency denoted by the vertical arrow, would change the absorption of the sample at that frequency, which can be used for storage of information. For example, a reduction of the absorption at the selected frequency would be denoted as a logic 1 and the full unchanged absorption can be denoted as logic 0. The total number of available frequency channels that can be stored is, therefore, $\Delta\nu_{in}/\Delta\nu_h$. The larger the $\Delta\nu_{in}$ and the smaller the $\Delta\nu_h$, the larger the storage density.

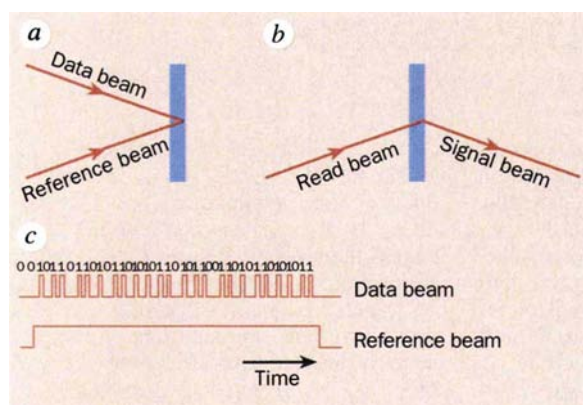
There are usually two approaches for spectrally addressing and storing the information. In the direct or frequency-domain approach, data are stored by directly tuning the laser to different frequencies inside the main absorption profile. On the other hand, in the time-domain approach, data are encoded in the temporal profile of the data beam, and the spectral channels are accessed in parallel. Here the Fourier spectrum of the data stream is recorded.

The work of Lin *et al.*¹ uses a hybrid of these two approaches, a process



referred to as swept-carrier time-domain approach. In this scheme, data are temporally encoded as in the time-domain approach. However, the data beam frequency is linearly changed (linear frequency chirp), as shown in part b of the figure. The reference beam is also linearly chirped. Among other conditions, the magnitude of the frequency chirp should be smaller or equal to the inhomogeneous linewidth to avoid exceeding the material frequency bandwidth. The swept-carrier storage is a superior approach, as it eliminates some of the drawbacks of both the direct and time-domain addressing. N. P.

- Price, L. C. & Schoell, M. *Nature* **378**, 368–371 (1995).
- Mango, F. D. *Geochim. cosmochim. Acta* (in the press).
- Mango, F. D., Hightower, J. W. & James, A. T. *Nature* **368**, 536–538 (1994).
- Helgeson, H. C. in *Geochemistry of Hydrothermal Ore Deposits* (ed. Barnes, H. L.) 568–610 (Wiley, New York, 1979).
- Steeffel, C. I. & Lasaga, A. C. *Am. J. Sci.* **294**, 529–592 (1994).
- Raffensperger, J. F. & Garven, G. *Am. J. Sci.* **295**, 639–696 (1995).
- Seewald, J. S. *Nature* **370**, 285–287 (1994).
- Shock, E. L. *Am. J. Sci.* **295**, 496–580 (1995).
- Helgeson, H. C., Knox, A. M., Owens, C. E. & Shock, E. L. *Geochim. cosmochim. Acta* **57**, 3295–3339 (1993).
- Huber, R., Kurr, M., Jannasch, H. W. & Stetter, K. O. *Nature* **342**, 833–834 (1989).



a, The data recording sequence with the reference and data beams illuminating the storage cell. b, The data retrieval sequence with a read beam illuminating the storage cell and the signal beam representing the retrieved data. c, Time profile of the reference and data pulses.

beam that is transmitted through the storage material. By moving the illuminating beam to a new spot on the material, a new storage cell is created, which can be used in a similar fashion for storage and retrieval of additional information. In this scheme, the information is recorded as a Fourier spectrum of the time-dependent data. The frequency of the data pulses is varied linearly ('chirped') and the spectral channels are accessed in parallel.

The medium used in this approach to data storage is typically an inhomogeneously broadened absorber, such as a crystal with an absorption profile consisting of many atomic-like transitions. The broader the inhomogeneous linewidth and the narrower the underlying individual transitions, referred to as the homogeneous linewidth, the larger the storage density. The particular material chosen by Lin *et al.* was a 500- μm -thick slab of a YAG (yttrium-aluminium-garnet) crystal doped with thulium (Tm^{3+}) atoms. The inhomogeneously broadened absorption profile of this crystal results from the transitions between the Tm^{3+} energy levels. This material was chosen because its absorption line at 793 nm (at the centre of the profile) is within the reach of commercially available semiconductor lasers. In the experiment, the data beam and the reference beam were focused to a spot size of $12\text{ }\mu\text{m} \times 16\text{ }\mu\text{m}$. A 1,760-bit-long sequence of binary data with a 50-ns bit duration was recorded in the illuminated region, corresponding to an areal storage density of 8 Gbits per square inch.

Laser tuning limitations have so far prevented the group making full use of the storage material bandwidth. Assuming future improvements in the laser tunability, the areal storage density can be extrapolated to about 100 Gbits per square inch, which would be a considerable improvement over the storage capacity of alternative approaches to optical recording.

To put this achievement in perspective,

IBM recently reported² a milestone in magnetic data storage in the laboratory, having reached an areal density of 3 Gbits per square inch. Magneto-optical storage technology has also produced 3 Gbits per square inch in the laboratory^{3,4}. Commercial magneto-optical disks are currently available with storage densities of 0.65 Gbits per square inch, and the latest commercially available magnetic disks used in personal computers reach densities of 0.2–0.3 Gbits per square inch. Holographic optical storage⁵, where a large number of high-

resolution images can be stored optically within the volume of a photorefractive material⁶, can also be used for parallel storage/retrieval of several million bits of information. As many as 5,000 holograms, each consisting of 320×220 pixels, have been stored using angle multiplexing in a 3 cm^3 photorefractive crystal⁵.

Of course, the new work is some way from practical application. For one thing, the sample had to be cooled to the cryogenic temperature of 4.8 K in order to reduce the homogeneous linewidth, thereby increasing the storage density. And the storage time, which is limited by the thermalization time of the ground-state sublevels, was a mere 200 μs . In general, long storage times of months or years are needed, although for some applications short storage times can still be useful. So, even though the results reported by Lin *et al.* are interesting, several obstacles have yet to be overcome in their approach. Storage materials that can operate at ambient temperatures around 300 K are needed, and longer data retention times are essential, so that the memory refresh process does not have to be repeated often. Nevertheless, with new materials and the performance leaps of the last few years, the spatial-spectral technique is very promising as a viable technique for the storage of information. $\square$

Nasser Peyghambarian is in the Optical Sciences Center and the Department of Materials Science and Engineering at the University of Arizona, Tucson, Arizona 85721, USA.

1. Lin, H., Wang, T. & Mossberg, T. W. *Opt. Lett.* **20**, 1658–1660 (1995).
2. Gains, B. *Condensed Matter News* **4**, 3–4 (1995).
3. Mansuripur, M. *The Physical Principles of Magneto-optical Recording* (Cambridge Univ. Press, London, 1995).
4. Gordon, G. R., Ooki, H. & Tyan, Y. S. *Proc. SPIE* **2514** (The International Society of Optical Engineers, Bellingham, Washington, 1995).
5. Mok, F. H. *Opt. Lett.* **18**, 915–917 (1993).
6. Meerholz, K., Volodin, B., Sandalphon, B., Kippelen, B. & Peyghambarian, N. *Nature* **371**, 497–500 (1994).

DAEDALUS

De-twinkled stars

THE fate of the Universe, or at least our opinion of it, depends on big astronomical telescopes. But as these optical marvels get ever bigger, their performance fails to improve as it should. The culprit is the atmosphere, whose convection and turbulence make light drift and wander. A big mirror receives light which has travelled by many different paths through the wavering atmosphere, giving blurred or multiple images. One answer is to avoid the densest air by mounting the telescope as high up as possible — the record to date being held by the orbiting Hubble Space Telescope. Daedalus once proposed to site big fans in front of a telescope, to stir the atmosphere forcibly into homogeneity. He now takes this idea a step further.

He points out that some microwave frequencies are strongly absorbed by air, notably by its content of water vapour. The absorbed energy appears as heat, and will set up convective flow in the warmed air. So to stir the atmosphere along the line of sight of a telescope, simply fire a suitably absorbed microwave beam out along it. Rapid pulsing and wagging of the beam could stir the air far faster and more thoroughly than any fan. Microwaves can be reflected from wire meshes that transmit light almost perfectly. A mesh mirror at 45° to the optical axis of the telescope could introduce a microwave beam into its light-path, but going the other way. It might even use the main mirror to focus it parallel. Fired out into the field of view, it would travel along the incoming light-path, stirring the precise column of air that causes the trouble. Experiment would soon establish the beam power and fluctuation pattern that gave the best seeing. The resolution of a 200-inch telescope might be improved from about 0.5 arcsec right up to the theoretical maximum of 0.02 arcsec.

But Daedalus goes even further. One anti-turbulence scheme employs rapid optical feedback. If the image wanders, fast piezoelectric actuators quickly tweak the telescope mirror to bring it back again. A feedback-controlled radiation beam could do a far better job. Instead of struggling to correct the problem by bending the mirror, it would eliminate it at source, by acting directly on the air causing the trouble. With no moving parts, it could react far faster and more exactly. Each photon of incoming light would be straightened out before it reached the telescope, and would contribute to a perfect and perfectly stable image. Black holes, missing mass, the enigmatic Hubble constant, all would be clarified by streams of refined and reliable data.

David Jones

Passing round the standards

SIR — Legislation to prevent the trade in endangered species enforced by the US Fish and Wildlife Service has alarmed biosystematists, as it may hinder the loan of collections vital for their research^{1,2}. Existing legislation, and administrative and policy information on access to genetic resources, was on the agenda of the Second Conference of the Parties (COP2) to the Convention of Biological Diversity in Jakarta, held over the past two weeks (6–17 November). If research is not to be frustrated and scientific names are to be applied accurately, recommendations must ensure that living and preserved scientific reference collections remain available to researchers world-wide. Name-bearing types in particular are the keystones of unequivocal communication in biology.

Types collected in less developed countries in the 1800s and early 1900s mostly reside in institutions of former colonial powers, or of countries from which expeditions were mounted. To ascertain current practice, we examined the countries of origin and place of deposit of the name-bearing type material ('holotypes') of fungi catalogued since 1991 (see table and ref. 3); no other world eukaryote nomenclatural index routinely collects these data.

The name-bearing types of 41% of the fungi described as new in this period were preserved in collections other than their country of origin. The situation was most acute for countries who are not members of the Organization for Economic Co-operation and Development (OECD), with 50% of the material outside, but remained significant for the OECD countries at 29%. In some non-OECD countries, the proportion was higher than in many OECD countries: Brazil, China, Cuba and India together retained the types of 1,116

(88%) fungi newly described from their territories, as opposed to only 147 (12%) deposited elsewhere. Even for OECD countries with major systematic institutions, a significant proportion of material was not deposited in them — 69 (17%) of 404 species described from the United States, and 15 (17%) of 90 species from the United Kingdom.

We conclude that, in at least one major group of organisms, a significant proportion of name-bearing types is still being deposited outside their country of origin. We also note that in some cases this practice will have been authorized as a part of collaborative programmes, and in others duplicates of the types ('isotypes') will have been deposited in the country of origin.

Tuning stochastic resonance

SIR — Collins *et al.*¹ claim to have removed a major obstacle to the use of stochastic resonance (SR) in biological and electronic systems. The 'tuning' of noise power, D , required for optimal signal detection in individual excitable units is said to become very broad when one sums the responses of a sufficient number, N , of independently noisy units driven by the same signal. This claim of "SR without tuning" is based on their equation (3), and illustrated by the flattening of the tuning curves in their Fig. 2. Throughout, a "normalized power norm" C_3 was used to measure output quality.

However, the flattening of these tuning curves for increasing N is due to the choice of C_3 and does not reflect any novel property of the SR model. The C_3 measure must saturate to 1, as its definition equals that of the standard correlation coefficient between input and (filtered) output. By contrast, tuning curves of constant shape are obtained when plotting the same results in terms of output signal-to-noise power ratio (SNR), or its logarithm. Under the same approximations as required for the validity of equation (3) of ref. 1 (additive, independent output noise), the SNR curves simply scale as $\text{SNR} \propto N$.

For periodic as well as aperiodic inputs, the relation of the two measures is $\text{SNR} = C_3^2/(1 - C_3^2)$, with its inverse $C_3^{-2} = 1 + \text{SNR}^{-1}$. Thus, C_3 saturates strongly in the regime $\text{SNR} \gg 1$ where $C_3 \geq 1/\sqrt{2}$, but behaves as $C_3 \approx \sqrt{\text{SNR}}$ in the lower regime. Combining this with $\text{SNR} \propto N$ fully explains the shape changes in the C_3 curves (Fig. 2 of ref. 1). Furthermore, it could well be that interesting features of the numerical data plotted in Collins *et al.*'s Fig. 3 are hidden in the highly saturated parts of the curves ($C_3 \geq 0.9$).

Against this background, it would not be inappropriate for the COP to refer the case of name-bearing type material for future consideration by its Subsidiary Body on Scientific, Technical and Technological Advice (SBSTTA). The explicit exemption of loans of all categories of permanently preserved name-bearing types from trans-national boundary controls would assist the biosystematists of all countries in the herculean task of naming the immense diversity of life on this planet.

David L. Hawksworth

Paul M. Kirk

*International Mycological Institute,
Bakeham Lane,
Egham TW20 9TY, UK*

1. Holmes, B. *New Scientist* **145**, issue 1960, 9 (1995).
2. Stone, R. *Science* **267**, 22 (1995).
3. Kirk, P. M. (ed.) *Index of Fungi* Vol. 6, pts 1–8 (CAB International, Wallingford, 1991–94).
4. *OECD Observer* **194**, June/July (1995).

For the applications suggested by Collins *et al.*, the SNR (or its logarithm) is not just a conventional measure, but also a natural one, reflecting the amount of information transmitted through the system. More precisely, Shannon's classic result² yields a transmission capacity (in bits per s) of $W \log_2(1 + \text{SNR})$, where W is the effective width of the (filtered) signal-spectrum. Thus, optimal information transfer through the system even shows a slight sharpening of its tuning curve with increasing N , contrary to the conclusions of Collins *et al.*

André J. Noest

*Ven 41,
NL-1115-HC Duivendrecht,
The Netherlands*

COLLINS ET AL. REPLY — Noest's result concerning the dependence of the output SNR on N is well known to us and others³. However, his statements concerning our Letter¹ are misleading.

As noted in our Letter, the normalized power norm C_3 saturates to 1 only when N equals ∞ . In that case, the output SNR equals ∞ over the same range of noise intensity, D . Thus, in the limit, the SNR curve 'flattens' (that is, its shape is not preserved). Clearly, our result does not depend upon whether one uses C_3 or the output SNR — the only difference between the two measures is in the normalization. This is also relevant to the biological and engineering applications we discussed. In such cases, the performance

LOCATIONS OF NAME-BEARING TYPES OF
NEWLY DESCRIBED
FUNGAL SPECIES, 1991–94

Country of origin	Type in-country	Type out-of-country	Total
OECD countries*	1,709 (30%)	710 (13%)	2,419 (43%)
Other countries	1,591 (29%)	1,563 (28%)	3,154 (57%)
Total	3,300 (59%)	2,273 (41%)	5,573 (100%)

*Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Japan, Luxembourg, Mexico, The Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Turkey, United Kingdom and United States⁴.

1. Collins, J. J., Chow, C. C. & Imhof, T. T. *Nature* **376**, 236–238 (1995).
2. Shannon, C. E. *Bell Syst. tech. J.* **27**, 379–423, 623–656 (1948).
3. Neuman, M. R. in *Medical Instrumentation: Application and Design* 2nd edn (ed. Webster, J. G.) 288–353 (Houghton Mifflin, Boston, 1992).
4. Collins, J. J., Chow, C. C. & Imhoff, T. T. *Phys. Rev. E* **52**, R3321–R3324 (1995).

of a system is optimally enhanced provided signal fidelity exceeds some level, which is set by the limitations of the system. We have demonstrated that, for a large enough N , any fidelity limit can be exceeded for a range of D . This applies equally to C_3 and the output SNR. For example, if for a range of D , C_3 exceeds 0.99, then similarly, Noest's output SNR will exceed a value of about 50 over the same range.

The output SNR is not a natural measure for the applications considered in our Letter, given that it is ill defined for systems with aperiodic inputs⁴ (Noest himself defines his output SNR in terms of C_3). The normalized power norm (or cross-correlation coefficient), on the other hand, can be directly determined for systems with aperiodic inputs and is thus an appropriate measure for characterizing the amount of information transmitted through such systems; it is widely used in neurophysiology for that purpose. Moreover, because the output SNR diverges as $N \rightarrow \infty$, we think that C_3 is a more intuitive measure of signal fidelity.

Finally, Noest uses Shannon's informa-

tion measure to argue for the use of the output SNR. We do not agree with his claim that the transmission-capacity tuning curve shows a slight sharpening with increasing N . If $G(D)$ represents the D -dependent output SNR of a single unit, for a system with N units the output SNR $\propto NG(D)$. Substituting this into the expression for the transmission capacity yields $W \log_2(1 + NG(D))$. For large N , this expression approximately equals $W[\log_2(N) + \log_2(G(D))]$. From this, one can clearly see that as N increases, the importance of the shape-dependent term decreases, contrary to Noest's statement.

Thus, for the model considered in our Letter, stochastic resonance remains without tuning, irrespective of the information measure used. This is consistent with both the title and conclusion of our Letter.

J. J. Collins

Carson C. Chow

Thomas T. Imhoff

NeuroMuscular Research Center,
Department of Biomedical Engineering,
Boston University,
Boston, Massachusetts 02215, USA

Toxic mutants in Charcot's sclerosis

SIR — Orrell *et al.*¹, while investigating the causes of familial amyotrophic lateral sclerosis (FALS, also known as Charcot's syndrome), have offered evidence, in red blood cells, for a mutation (G93R) in superoxide dismutase 1 (SOD1) that reduces the enzymatic activity of wild-type SOD1 in this autosomal dominant disease. As disease onset in this G93R family is around 36 years, compared with an average of 50 years for other mutants, it was proposed that dominant loss of activity underlies disease for this mutant. This view reinforced the initial proposal that disease arises from loss of activity².

Whether loss of SOD1 activity, and by inference oxidative stress/damage, mediates selective motor neuron death in this disease is a crucial question for designing therapeutic approaches. There is, however, considerable uncertainty as to whether activity levels in biosynthetically inactive red blood cells are appropriate indicators

of SOD1 levels in the motor neurons at risk. This ambiguity is compounded by the absence of evidence that G93R, or other FALS mutants, reduce wild-type SOD1 activity in cell types that, like motor neurons, are translationally active.

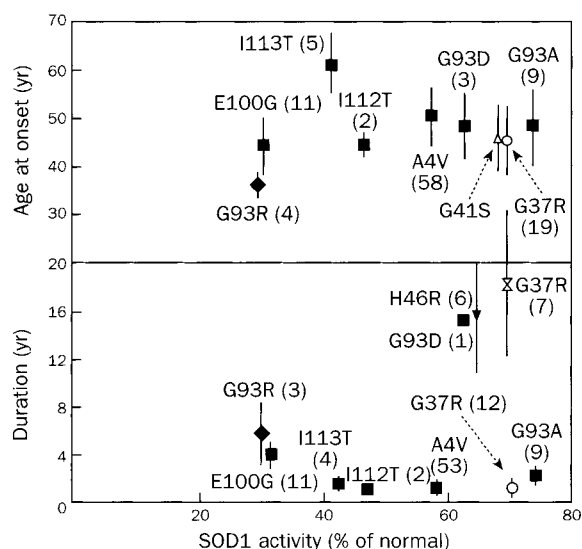
Even without these confounding influences, consideration of activity levels measured in extracts of either red blood cells or lymphoblasts of patients now available, representing ten different point mutants in SOD1, does not reveal a convincing correlation between activity and age of disease onset or length of survival after onset (see figure). Moreover, some mutations yield only small reductions in activity, despite the possibility that the activity of some mutants (■ and ▼ in the figure) may actually be underestimated (due to assaying after the extraction of haemoglobin). In particular, two mutants, G93A and G37R, cause disease despite less than 30% reductions in normal SOD1 activity in patient samples^{3,4}. For G37R, a similarly small reduction is seen in comparing activities in monkey fibroblasts

expressing human wild-type or G37R SOD1 (ref. 4). Moreover, G37R retains full intrinsic specific activity⁴, and its expression causes motor neuron disease in mice while raising overall SOD1 activity in the spinal cord 5–12-fold⁵. G37R mediates disease in humans with an average onset of 46 years ($n=19$). This is similar to that of mutants with significantly lower activities.

A final feature that strongly suggests the involvement of factors beyond SOD activity is that, although the onset of G37R-mediated disease is similar in Australian and North American families, disease progresses very rapidly in the Australian patients (1 yr duration, $n=12$), but much more slowly in the North American family (18 yr duration, $n=7$).

Further evidence against loss of activity *per se* as the crucial property that results in FALS is the conspicuous absence, despite around 30 known point mutations in SOD1, of lesions that block the synthesis of a full-length or nearly full-length SOD1 polypeptide. Moreover, at least three mutants can confer motor neuron disease in mice without reducing SOD1 activity^{5–7}. For each of these mice, disease is most readily explained by gain of a deleterious property by the mutant subunits.

A unifying view would be that, although most mutants lead to reduced SOD1 activity, and some (like the G93R mutant reported by Orrell *et al.*¹ and several



Average age of disease onset or duration in patients with FALS caused by 10 different mutations in SOD1 plotted against SOD1 activity measured in lysates of red blood cells or lymphocytes. Vertical lines indicate standard deviations. Numbers in parentheses denote number of patients for which onset or disease duration are available. Activity measurements from red blood cells (filled symbols) and lymphoblasts (open symbols) were taken from ref. 1 (◆), ref. 3 (■), ref. 4 (◻), ref. 9 (△) and ref. 10 (▼). Duration of disease mediated by G37R is represented by two points, one for Australian (○) and one for North American patients (◻). ■, ▼, Assayed following removal of haemoglobin by organic extraction; ◆, △, assayed in solution without organic extraction; ◻, ○, assayed at neutral pH on nondenaturing gels.

- Orrell, R., Belleruche, J., Marklund, S., Bowe, F. & Halliwell, R. *Nature* **374**, 504–505 (1995).
- Deng, H. X. *et al. Science* **261**, 1047–1051 (1993).
- Bowling, A. C. *et al. J. Neurochem.* **64**, 2366–2369 (1995).
- Borchelt, D. R. *et al. Proc. natn. Acad. Sci. U.S.A.* **91**, 8292–8296 (1994).
- Wong, P. C. *et al. Neuron* **14**, 1105–1116 (1995).
- Gurney, M. E. *et al. Science* **264**, 1772–1775 (1994).
- Ripps, M. E. *et al. Proc. natn. Acad. Sci. U.S.A.* **92**, 689–693 (1995).
- Phillips, J. P. *et al. Proc. natn. Acad. Sci. U.S.A.* **92**, 8574–8578 (1995).
- Tsuda, T. *et al. Neuron* **13**, 727–736 (1994).
- Ogasawara, M. *et al. Nature Genet.* **5**, 323–324 (1993).
- Dal Canto, M. C. & Gurney, M. E. *Brain Res.* **676**, 25–40 (1995).
- Andersen, P. M. *et al. Nature Genet.* **10**, 61–66 (1995).
- Rosen, D. R. *et al. Hum. molec. Genet.* **3**, 981–987 (1994).

chemically induced mutants in *Drosophila*⁸) partially inactivate or destabilize the wild-type subunits in some cells, loss of activity in FALS is coincidental to the acquisition by the mutant subunit of an as yet unidentified toxic property that mediates selective motor neuron death.

Don W. Cleveland

Ludwig Institute for Cancer Research
and Departments of Medicine and
Neuroscience,
University of California at San Diego,
La Jolla, California 92093, USA

Nigel Laing

Patricia V. Hurse
Molecular Neurogenetics Laboratory,
Queen Elizabeth II Medical Center,
Nedlands, Western Australia 6009,
Australia

Robert H. Brown Jr

Day Laboratory for Neuromuscular
Research,
Massachusetts General Hospital,
Charlestown, Massachusetts 02129, USA

ORRELL *ET AL.* REPLY — First, we agree with Cleveland *et al.* that SOD1 FALS is caused by a toxic property present in CuZn superoxide dismutase (SOD) which is uncovered by the mutations. In our view¹, the best evidence supporting this conclusion comes from experiments with transgenic mice overexpressing various SOD1 FALS proteins. These animals develop a FALS-like disease, yet have greatly increased levels of SOD activity in central nervous system tissue^{5,6,11}. Some other studies, such as those on the D90A mutant which has essentially normal activity in erythrocytes, also support the gain of a toxic function¹². Research in SOD1 FALS now centres on determining the nature of the toxicity — its molecular determinants in SOD, the pathway(s) for transmission of toxicity and the cellular target(s).

Such studies may lead to rational approaches to therapy. We do not believe, however, that the data presented in the figure can be used to assess whether loss of SOD activity is involved in the disease process. Apart from the G37R and G41S mutants, the enzyme activity measurements have been carried out on erythrocyte lysates subjected to organic extraction to remove haemoglobin which would otherwise interfere with the assay used. The consequence of this harsh extraction procedure and subsequent centrifugation is variably to precipitate and remove homodimers and heterodimers containing the less stable mutant subunits¹ (SOD is a homodimer and affected individuals are heterozygotes). We believe this invalidates the use of these erythrocyte data as an attempt to correlate activity with onset or duration. The remaining two mutants (G37R and G41S) were assayed in lymphoblastoid cells and, of these, G41S is

from a single living individual and cannot be statistically meaningful.

We do not know, at present, whether the much reduced SOD activity found in erythrocytes containing G93R is an important or merely an incidental property of this mutant. Nonetheless, this mutant is of considerable interest because it is the only characterized SOD1 FALS mutant that appears to cause early disease onset and has less than 50% of the enzyme activity found in control erythrocytes. The average age of onset for FALS is 49.1 ± 13.7 yr (mean $\pm$ s.d.), and average duration is 2.5 ± 1.9 yr¹³, whereas for the G93R family, onset averages 35.8 ± 4.3 yr and duration averages 5.7 ± 4.5 yr¹. There seem to be two, at present indistinguishable, mechanistic possibilities for G93R: that reduced enzyme activity is contributing to early disease onset, or that G93R is more toxic than other mutants and the reduced activity is incidental.

Investigation of mutations that alter the normal age of onset or duration of the disease may provide insight into the nature of the disease mechanism. In addition to G93R, A4V has normal onset but short

duration (1.2 ± 0.8 yr)¹³, and H46R appears to have normal onset and long duration (17.3 ± 10.8 yr)¹⁰. If disease severity was a simple function of SOD toxicity, more toxic mutants would be expected to have early onset and short duration. The finding that onset and duration may be independently altered by particular mutations suggests that another property of the mutant SODs, in addition to the proposed toxicity, is involved in the disease mechanism.

Richard W. Orrell

Jaqueline de Bellerocche

Department of Biochemistry,
Charing Cross and Westminster
Medical School,
Fulham Palace Road, London W6 8RF, UK

Stefan L. Marklund

Department of Clinical Chemistry,
Umeå University Hospital,
S-901 85 Umeå, Sweden

Frances A. Bowe

Robert A. Hallelwell

Department of Biochemistry,
Imperial College of Science,
Technology and Medicine,
Exhibition Road, London SW7 2AY, UK

Origin of yaws in the Pleistocene

SIR — Although controversy exists about the role of Columbus in the spread of syphilis^{1–4}, new observations now suggest that treponematoses (in the form of yaws) actually had its origin in Africa during the Middle Pleistocene (1.5 million years ago). This conclusion is based on reanalysis of periosteal reaction patterns in *Homo erectus* KNM-ER 1808 and study of other *H. erectus* skeletal remains in the National Museum of Kenya in Nairobi.

Previous attempts to determine the origin of treponemal disease have been complicated by an inability to distinguish microbiologically among treponemoses⁵. The availability of criteria for distinguishing between syphilis, yaws and bejel (nonvenereal syphilis)⁶ allows a new perspective. Osseous reactions to treponemal infection, although reproducible for each variety, are not uniform among them⁶. Examination of population frequency, demographics, character and skeletal distribution, however, provides clear, reproducible clues to the presence of treponematoses and to specific identity of the underlying bone-attacking treponematoses⁶.

We examined skeletal remains of *H. erectus* from Koobi Fora Formation, East Lake Turkana in Kenya (dated at 1.6 ± 0.1 million years)⁷. Humeral fragments from 13 individuals labelled as *H. erectus* and tibial fragments from two individuals were present, in addition to the more complete skeletons of KNM-WT 15,000 and KNM-ER 1808. Within the preserved postcranial skeleton in KNM-ER 1808 (humeri, radii, ulnae, femora, tibiae, fibula and clavicle), greatly thickened diaphyses (periosteal reaction up to 7 mm thick) were noted, sparing only the clavicle. Periosteal reaction was also noted in the isolated humerus, KNM-ER 737. The cross-sectional appearance of KNM-ER 1808 was assessed at the site of a post-mortem break in an affected fragment. Sharply demarcated new periosteal bone was noted, identical to that seen in contemporary individuals with treponematoses⁸. Histological preparation was, unfortunately, not permitted.

Although Walker *et al.*⁷ have suggested a diagnosis of hypervitaminosis A, the absence of enthesial reaction⁹ makes this difficult to substantiate. Humeral sparing,

1. Baker, B. J. & Armelagos, G. J. *Curr. Anthropol.* **29**, 703–720 (1988).
2. Hudson, E. H. *Br. J. vener. Dis.* **34**, 22–24 (1958).
3. Hackett, C. J. *Bull. WHO* **29**, 7 (1963).
4. Cockburn, A. *Infectious Diseases: Their Evolution and Eradication*, 48–49 (Thomas, Springfield, IL, 1967).
5. Stewart, T. D. & Spoehr, A. *Bull. Hist. Med.* **26**, 538–541 (1952).
6. Rothschild, B. M. & Rothschild, C. *Clin. Infect. Dis.* **20**, 1402–1408 (1995).
7. Walker, A., Zimmerman, M. R. & Leakey, R. E. *Nature* **296**, 248–250 (1982).

8. Hershkovitz, I., Rothschild, B. M., Wish-Baratz, S. & Rothschild, C. in *The Origin of Syphilis in Europe* (eds Dutour, O., Palfi, G., Berato J. & Brun, J.-P.) 81–87 (Centre Archéologique du Var, Toulon, 1995).
9. Seawright, A. A., English, P. B. & Gartner, R. J. *J. comp. Path.* **77**, 29–43 (1967).
10. Resnick, D. & Niwayama, G. *Diagnosis of Bone and Joint Disorders* (Saunders, Philadelphia, 1989).
11. Rothschild, B. M. & Martin, L. *Paleopathology: Disease in the Fossil Record* (CRC Press, London, 1993).
12. Fornaciari, G. *et al. J. comp. hum. Biol.* **45**, S49 (1995).

its extreme rarity in adults and complete resolution with cessation of intake of excess vitamin A make hypervitaminosis A an untenable diagnosis⁸.

Among those disorders producing periosteal reactions, a polyostotic disease such as seen here is rare, with involvement of long bones of both upper and lower extremities reported only in hypertrophic osteoarthropathy and the treponemal disease yaws^{6,10,11}. Syphilis and bejel are predominantly pauci-ostotic diseases⁵. Humeral involvement in hypertrophic osteoarthropathy is extraordinarily rare and quite mild^{10,11}, contrasting with the significant thickening noted in KNM 1808. The distribution pattern of periosteal reaction in KNM 1808 therefore appears most compatible with a mainly polyostotic yaws, typically affecting all long bones.

These findings seem to confirm Cockburn's¹ and Hudson's² hypothesis of an African origin for treponemal disease. It sets the date of origin to the Middle Pleistocene, considerably earlier than previously proposed. The presence of such periostitis has been reported in an isolated *H. erectus* femur, aged at 500,000 years before present, from Venosa (Basilicata, southern Italy)¹². This further supports the contention that *H. erectus* carried the disease to Europe.

Bruce M. Rothschild*

Israel Hershkovitz†

Christine Rothschild*

*Arthritis Center of Northeast Ohio,
Youngstown, Ohio 44512, USA

†Sackler School of Medicine,
Tel Aviv University,
Ramat Aviv 69978, Israel

Legume nitrogen fixation and drought

SIR—Grain legumes are critically important in many cropping systems because of their ability to fix atmospheric nitrogen, and this importance is expected to increase with the need to develop sustainable agricultural practices. Unfortunately, N₂ fixation has been found to be sensitive to modest soil drying in soybean (*Glycine max* L. Merr.)¹, cowpea (*Vigna*

unguiculata L.)² and black gram (*Vigna mungo* L.)³, resulting in decreased yielding capacity. Here we show that drought sensitivity is not universal among legumes. Grain legumes that transport fixed nitrogen as amides have N₂-fixation activity that is less sensitive to soil drying than those that transport ureides, making them advantageous in dryland regions.

We tested chickpea (*Cicer arietinum* L.), common bean (*Phaseolus vulgaris* L.), faba bean (*Vicia faba* L.), lupin (*Lupinus albus* L.), pea (*Pisum sativum* L.) and peanut (*Arachis hypogaea* L.) for sensitivity of N₂ fixation to drought. These tests used an *in situ* assay requiring less than 15 minutes' exposure to acetylene³ to overcome problems resulting from disturbance of the plants⁴ and long-term exposure to acetylene in the assay⁵. Plants were grown in 2.3-litre pots for 4–6 weeks and then subjected to a dehydration cycle lasting about 2 weeks. The pots were sealed each afternoon and a 1:9 acetylene:air mixture was continuously flowed at 1 litre per min through the pots. After 10 min equilibration, gas coming from the pots was sampled and the ethylene concentration measured. The relative N₂-fixation activity (RDN) was obtained as the ratio between individual water-deficit pots and the mean of

well-watered pots. In addition, the pots were weighed each afternoon to calculate the relative transpiration rate (RTR) for each water-deficit pot.

The sensitivity of N₂ fixation to water deficits relative to the sensitivity of transpiration rate was resolved by calculating the ratio of RDN to RTR for each plant on each day during the dehydration cycle³. Ratios of less than one indicated that N₂ fixation had greater sensitivity to water deficits than did transpiration. The data obtained with soybean and cowpea showed that the sensitivity of RDN to soil drying in these species was less than that of RTR in nearly all cases. Surprisingly, all other grain legumes showed the previously unobserved phenomenon that RDN was less sensitive to water deficits than was RTR during the water-deficit period (see table). Grain legume species can, therefore, be selected for decreased sensitivity of N₂ fixation to soil drying for regions where drought is a recurring problem.

Segregation of legume species based on N₂-fixation sensitivity to drought showed a correlation with the type of nitrogenous compounds exported from nodules. Those species that transported high concentrations of ureides^{6,7} were found to be drought-sensitive. In our studies, the drought sensitivity of soybean and cowpea was associated with the highest xylem sap ureide concentrations (see table). Species with low or no ureide were found to be drought-tolerant. The association of high ureide concentrations in species with drought sensitivity may result from the low solubility of ureides⁶. Decreased xylem transport in the ureide-transporting species with soil drying might result in a negative feedback on N₂ fixation⁸, and consequently result in the expressed drought sensitivity of several important grain legume crops.

T. R. Sinclair

USDA-ARS, Agronomy Department,
Agronomy Physiology Laboratory,
University of Florida, PO Box 110840,
Gainesville, Florida 32611, USA

R. Serraj

Laboratoire de Physiologie Végétale,
Département de Biologie,
Faculté des Sciences-Semlalia,
BP S 15 Marrakech, Morocco

1. Djekoun, A. & Planchon, C. *Agrochim.* **83**, 316–322 (1991).
2. Sinclair, T. R. *et al. Field Crops Res.* **17**, 121–140 (1987).
3. Sall, K. & Sinclair, T. R. *Pl. Soil* **133**, 31–37 (1991).
4. Sung, L., Moloney, A. H., Hunt, S. & Layzell, D. B. *Physiol. Pl.* **83**, 67–74 (1991).
5. Minchin, F. R., Witty, J. F., Sheehy, J. E. & Muller, M. J. *exp. Bot.* **34**, 641–649 (1983).
6. Sprent, J. I. *Pl. Cell Envir.* **3**, 35–43 (1980).
7. Atkins, C. A. in *Biology and Biochemistry of Nitrogen Fixation* (eds Dilworth, M. & Glenn, A.) (Elsevier, Amsterdam, 1991).
8. Parsons, R., Stanforth, A., Raven, J. A. & Sprent, J. I. *Pl. Cell Envir.* **16**, 125–136 (1993).
9. Young, E. G. & Conway, C. F. *J. biol. Chem.* **142**, 839–853 (1942).
10. Pate, J. S., Atkins, C. A., White, S. T., Rainbird, R. M. & Woo, K. C. *Pl. Physiol.* **65**, 961–965 (1980).
11. Peoples, M. B., Pate, J. S., Atkins, C. A. & Bergersen, F. J. *Pl. Physiol.* **82**, 946–951 (1986).

SPECIES COMPARISON OF SENSITIVITY OF N₂ FIXATION AND TRANSPIRATION TO SOIL DEHYDRATION

Species	No. of RDN observations	Fraction with RDN:RTR >1.0	Sap ureide (μM)*
Soybean cv. Biloxi	51	0.10	458
Cowpea ² cv. CB9	21	0.14	216
Black gram ² cv. Regur	10	0.30	†
Common bean cv. Roma	29	0.86	45
Pea cv. sugar snap	41	0.95	15
Lupine cv. Ultra	44	0.77	0
Peanut cv. Florunner	80	0.94	‡
Faba bean cv. F223	23	0.96	0
Chickpea cv. Sombrero INRA 199	37	0.97	0

Comparison of fraction of daily observations on individual plants where relative dinitrogen fixation is less sensitive to soil dehydration than is relative transpiration rate, and the xylem sap ureide concentration of unstressed plants.

*All species were grown simultaneously in a greenhouse for 5 weeks. Xylem sap was extracted by detaching the shoots, placing them in a pressure chamber and subjecting the shoot to a 1-MPa overpressure. The sap was immediately frozen and stored at –20 °C until ureide concentration was assayed colorimetrically⁹.

†Pate *et al.*¹⁰ found high levels of ureides in xylem sap of black gram.

‡HPLC analysis of peanut xylem sap has failed to detect the presence of ureides¹¹.

The artful scientist

L. Pearce Williams

Science and Culture: Popular and Philosophical Essays. By Hermann von Helmholtz, edited by David Cahan. University of Chicago Press: 1995. Pp. 418. \$52, \$41.50 (hbk); \$17.95, £14.25 (pbk).

HERMANN Helmholtz (the von was an honour celebrating his later scientific achievements) was born in Potsdam in Prussia in 1821 and died in Berlin, the capital of the new German Empire, in 1894. His father was a poor school-teacher, passionately fond of the new philosophies in Germany that followed on from the writings of Immanuel Kant. He also loved music and art. These passions were passed on to his eldest child and provided Helmholtz with a lifetime of both enjoyment and stimulus for much of his scientific work. As an adult, he was a piano player of concert quality and a sensitive participant in the art of the nineteenth century. Much of his scientific work was devoted to the scientific basis of musical harmony and the aesthetic appreciation of art. In the local gymnasium, Helmholtz received an excellent classical education, mastering both Greek and Latin and enjoying particularly the poetry of those two cultures. Music, painting and poetry were never far from his mind, even when working on the most abstruse subjects.

Unable to afford a university education, Helmholtz entered medical school in Berlin with the financial backing of the government, in return for which he was obliged to serve as an army doctor for eight years. It was in medical school that Helmholtz learned the fundamentals of anatomy and of accurate anatomical observations. It was also in Berlin that he was exposed to the medical theories of the day, which he found grossly unsatisfactory, and the physiological research of Johannes Müller, which was to lead to his lifelong interest in epistemology and neurology. A few courses in chemistry and his own study of mathematics and the philosophy of Kant completed his formal education.

This short sketch can only hint at the manifold genius of Helmholtz. His entry in the *Dictionary of Scientific Biography* lists his eminence in the following fields: energetics, physiological acoustics, physiological optics, epistemology, hydrodynamics and electrodynamics, without mentioning his work in non-Euclidean geometry. This, I believe, is the longest list of "specialties" to be found in the *DSB*. With the exception of electrodynamics, all these subjects are represented in his *Popular and Philosophical Essays* together with an autobiographical sketch near the end.

It is a small point, but I would have preferred this sketch at the beginning, for it provides, in short compass, a map of Helmholtz's career.

Despite the variety of topics treated, there are only a few fundamental ideas that guided Helmholtz's thought. The most important was the idea of scientific laws. Time and again Helmholtz refers to



Helmholtz: manifold genius guided by few fundamental ideas.

the fact that he found it difficult to deal with discrete facts (probably the reason chemistry does not figure centrally in his thought). Laws, he insisted, were established by induction from sensory experience and deduction, through mathematics, from empirical generalizations. His interest in philosophy led him to cite Kant, time and time again, as an empiricist — a position most university students of today might find a bit bizarre.

Given the centrality of the senses, the mechanism of sensory perception was of fundamental importance to him. Early in his career, he had worked in the laboratory of Johannes Müller, who had discovered the law of specific nerve energies: in other words, that each sensory nerve 'reported' that sense no matter what the source of stimulation. Thus the belief that seeing was believing could no longer be held. It was important, therefore, to look into the physiology of sensation, and this Helmholtz did. He invented the ophthalmoscope and the ophthalmometer that

permitted investigation of the eye. The ear posed anatomical problems that he was able to clarify, if not solve. These subjects are treated in four of the essays. Sensory physiology is also applied to the problem of aesthetic appreciation. As a superb musician, he wrote with authority in the essay entitled "On the physiological causes of Harmony in Music" and summed up his work in this area in "The Facts of Perception".

His belief that all of nature was subject to natural laws permitted him to dismiss the older systems of medicine and insist on scientific investigations of medical phenomena. "On Thought in Medicine" nicely describes the contemporary medical world of medicine which, after millennia of magic and metaphysical vapourings, began to be erected on a truly scientific foundation. Nor could he believe in the idea of perpetual motion, for that violated all the laws of nature as he knew them. His work in physiology and theoretical mechanics convinced him that 'force' (only later did he adopt the term 'energy') could only be transformed, not destroyed. This was, incidentally, an idea that had inspired a previous generation of German-speaking philosophers (known as *Naturphilosophen*) to insist that all the forces of nature were convertible into one another. Helmholtz was perfectly aware of this intellectual stream, although he deplored its metaphysical excesses. What had to be done was to derive this law from empirical and mathematical principles. This he did in 1847 (at the age of 26) along with others with whom he shares the distinction. His paper "Über die Erhaltung der Kraft" made a deep impression on his contemporaries and its mathematical message is here simplified in "On the Conservation of Force".

Finally, there are articles on Goethe, who fascinated Helmholtz, possibly because Goethe's theory of colours was, in his opinion, so profoundly wrong. The essays "On Goethe's Scientific Researches" and "Goethe's Presentiments of Coming Scientific Ideas" explore with great sensitivity the difference between the poet and the scientist but, instead of describing two cultures deaf and blind to each other, Helmholtz tries, with some success, to show how they are complementary.

The volume is handsomely presented, although I am mystified by the cover depicting a flute concert with the flute player wearing an eighteenth-century wig and jackboots (Frederick the Great?). The contents, however, are fascinating. □

L. Pearce Williams is in the Department of History, Cornell University, Ithaca, New York 14853, USA.

Algal aesthetics

G. E. Fogg

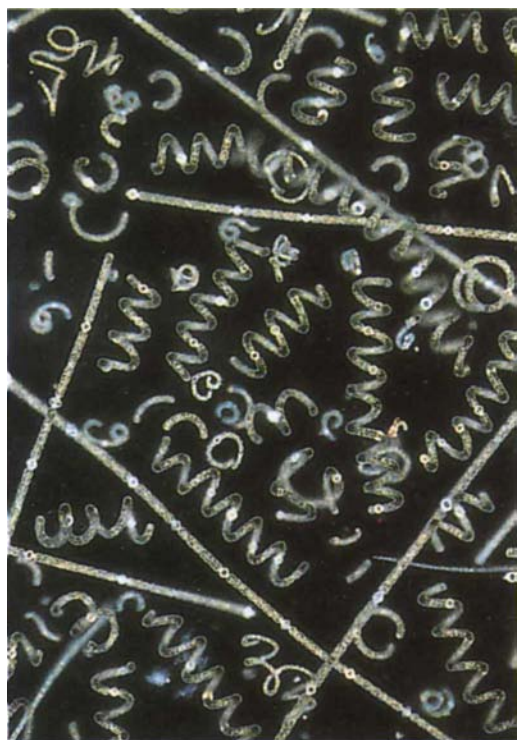
Freshwater Algae: Their Microscopic World Explored. By Hilda Canter-Lund and John W. G. Lund. Biopress: 1995. Pp. 360. £49.50.

THE gulf between the two cultures is particularly evident where algae are concerned. Since Virgil wrote the words *nihil vilior alga* — there is nothing worse than an alga — literary persons have done little more than mention them as bloom-inspiring items in the scenery. Frequent reports in the media of algal blooms causing outbreaks of poisoning have done nothing to improve this image. Whereas the flourishing art of botanical illustration has bridged the gulf with pictures of flowering plants that give both satisfaction to the scientist and delight to connoisseurs of beautiful drawing, illustrations of algae have not attracted similar aesthetic appreciation. Nevertheless, as the layman who looks at this book will readily understand, not the least of the reasons for biologists taking up the study of these microscopic creatures is their beauty and variety of form.

The basis of the book is a collection of photomicrographs of freshwater algae, mostly from the English Lake District, taken by Hilda Canter-Lund, who is not only an expert on the fungi and protozoa that attack algae but also a photographer of genius. The photographs have been taken with serious scientific purpose, using living material and, as necessary, dark-field illumination, phase-contrast or Nomarski differential interference contrast to clarify structural features. There

is no obvious striving to create artistic arrangements, but, no doubt by exercise of infinite patience, the camera has been used creatively to dispose the objects satisfyingly within the frame and to display their jewel-like colours and shapes to best advantage. A designer of fabrics might well be inspired by the golden-brown rosettes of *Synura*, the *Ceratium*-dominated plankton looking like a scattering of mediaeval weapons, or the pattern of straight and spiral *Anabaena* filaments. A dramatic mass attack of the amoeboid *Vampyrella* is shown. Altogether, there are 387 colour and 253 monochrome photographs in the book, all of first-class quality. The photographer will always have some criticism of reproductions of his or her colour photographs but, to one who is familiar with the algae depicted, the colour balance seems uniformly good. Only very occasionally is one unconvinced by the pictures of the existence of the structures described in the text.

The text is written mainly by the photographer's husband, J. W. G. Lund, a distinguished authority on freshwater algae. It started out as a commentary on the pictures but, without getting out of hand, developed into an admirable introduction to phycology. Although it is a continuous narrative, it has been skilfully designed so that the relevant text is, nearly always, on the page opposite the photograph. The style is simple and clear without sacrificing scientific precision, and, if one is in doubt, there is a glossary,



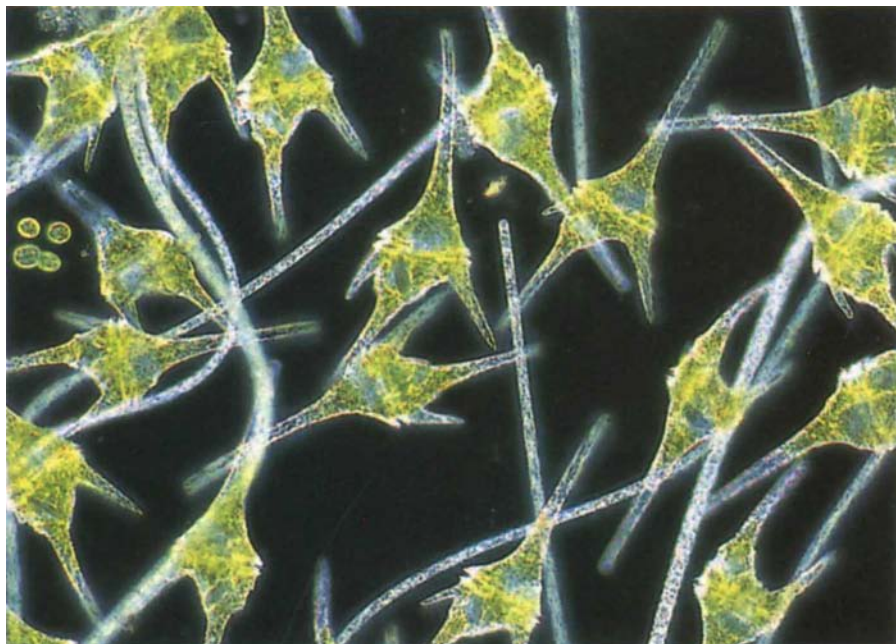
From Fresh Water Algae

Straight and spiral *Anabaena* filaments (width $\sim 10 \mu\text{m}$)

together with suggestions for further reading and separate indices of genera of algae, fungi and animals. The algal groups, with which are included the cyanobacteria, are dealt with in turn with accounts of structure, reproduction, occurrence and ecology of the species depicted. Attention is paid to the factors controlling the growth of phytoplankton, the problems of management of water bodies and eutrophication. The final chapters are concerned with symbiosis, animals feeding on algae and fungi living on and in algae. These are particularly valuable because they bring together material that is scarcely dealt with in the textbooks — snippets of natural history some might think, but essential information if one is to understand the wax and wane of populations, the succession of species and the place of algae in ecosystems.

This beautifully produced book has been put together with the object of interesting the general reader in the algae. Certainly it will stand out on the coffee-table and captivate those with an artistic eye, but it ought also to be accessible to every sixth-former and undergraduate studying biology. To professional phycologists and freshwater biologists it will bring a resurgence of the excitement they felt when they first encountered the microscopic algae, as well as a great deal of new and useful information. □

G. E. Fogg, emeritus professor of marine biology at the University of Wales, is at Bodolben, Llandegfan, Anglesey, Gwynedd LL59 5TA, UK



From Fresh Water Algae

Ceratium-dominated plankton (cell length $\sim 170\text{--}200 \mu\text{m}$).

Conservation biology defined

Katherine Ralls

A Primer of Conservation Biology. By Richard B. Primack. *Sinauer*: 1995. Pp. 277. £14.95 (pbk).

Principles of Conservation Biology. By Gary K. Meffe and C. Ronald Carroll. *Sinauer*: 1995. Pp. 600. £29.95.

Conservation Biology in Theory and Practice. By Graeme Caughley and Anne Gunn. *Blackwell Science*: 1995. Pp. 459. £32.50, \$45.95 (pbk).

Encyclopedia of Environmental Biology. Edited by William A. Nierenberg. *Academic*: 1995. Pp. 2,114 (3 volumes). \$475, £295.

A NOVICE student stumbling across one of these three texts on conservation biology — the study and protection of biodiversity — would get one of two very different views of this young scientific field. The first two books, by Richard Primack and by Gary Meffe and Ronald Carroll, cover biological diversity at all levels from genetic to global and include material from environmental ethics to the economics of conserving species and ecosystems. By contrast, the book by Graeme Caughley and Anne Gunn would lead our student to believe that conservation biology mainly consists of efforts to save endangered birds and mammals. And if our student found only the fourth book, an encyclopaedia, she would gather no inkling of the existence of conservation biology, as the topic is not addressed.

Burgeoning science

The growth of conservation biology in the United States, where the field was founded, has been extraordinary. The first conference on the subject was held in San Diego in 1978. At the second, in Michigan in 1985, participants voted to establish the Society for Conservation Biology. The first issue of the society's journal, *Conservation Biology*, appeared in May 1987. Despite the lack of a textbook until the fall of 1993 (when Primack's earlier book *Essentials of Conservation Biology* was published), undergraduate courses in conservation biology proliferated across the country, and by 1995 at least 51 US universities were offering graduate programmes in the field.

The texts by Primack and by Meffe and Carroll present the mainstream view in the United States that conservation biology is a broad response by scientists and other academics to the environmental devastation caused by ever-increasing numbers of human beings. In this view,

conservation biology is based on the belief that the diversity of organisms is intrinsically good, and the field has the mission of preserving natural ecosystems and biological processes.

A Primer of Conservation Biology is a shorter version of Primack's earlier book and portrays conservation biology as a rapidly changing, multidisciplinary, crisis-driven discipline. It describes biological diversity at the genetic, population, species and community levels and stresses the need to conserve diversity on all levels. It is well-suited for its purpose: to convey the essentials of the field to a wider audience in "general biology, ecology, wildlife biology, and environmental policy courses" and professionals in other disciplines. The final chapter, on conservation and sustainable development, outlines major problems and appropriate responses and encourages the reader to take action; it is followed by a helpful appendix listing selected environmental organizations and sources of information.

Principles of Conservation Biology covers much the same ground but in far greater detail. A comprehensive volume intended for advanced undergraduate and graduate students, Meffe and Carroll's text is complementary to Primack's books. It is a large and impressive volume, noteworthy for its emphasis on values, ethics and the need for individual action. For instance, the second chapter is devoted to conservation values and ethics, covering the instrumental and intrinsic approaches to valuing biodiversity and the bases of anthropocentrism, the Judaeo-Christian stewardship conservation ethic, traditional non-Western environmental ethics, biocentrism and ecocentrism.

In their final chapter, the authors return to the question of values, stressing the need to adjust human value systems to reflect ecological reality. The Earth cannot indefinitely withstand the cumulative impacts of a large and rapidly increasing human population (currently about 5.5 billion people, increasing by 95 million per year). Each chapter includes discussion questions designed to make the reader think about practical applications of the material presented, and the book's concluding sentence is: "Where do you think you could make the best contributions, given your own talents, interests, and experiences?"

Meffe and Carroll enlisted the support of their colleagues because they realized that two people "cannot comprehensively and fairly represent the broad interests of conservation biology without input from others". Thanks to their good taste in colleagues, enlisting others proved to be a splendid idea. The multiple and sometimes conflicting viewpoints expressed by these outside contributors

successfully convey the diversity of the conservation biology enterprise to the reader. Colleagues contribute all or part of 9 of the 18 chapters, as well 11 case studies and more than 50 essays. The essays are a diverse and fascinating group that includes "Discovering radical environmentalism in our own cultural backyard", "Valuation of extractive medicines in tropical forests" and "Pupfish, politics, and preserve management in the arid Southwest".

Rather than inviting colleagues to contribute supplementary material, Caughley and Gunn chose to restrict *Conservation Biology in Theory and Practice* to the areas they know best. Unfortunately, the result is a text too narrow for most US courses on conservation biology. In this book, conservation biology consists almost entirely of efforts to prevent the extinction of individual populations and species of higher vertebrates that are already in dire straits. Plants are not dealt with and the preservation of communities, landscapes and systems is given short shrift — a single chapter entitled "Reserves in theory and practice".

Limited data

Caughley and Gunn stress the need to diagnose the reasons for the decline of a species, using standard statistical testing of hypotheses, before taking action aimed at reversing the decline. Would that we had the time, funding and skill to implement this advice consistently. Unfortunately, at least in the United States, legal directives to base decisions on the best possible scientific data are not accompanied by requirements to collect even minimal data about threatened vertebrates, much less the myriad threatened invertebrates and plants. All too often, conservation biologists must choose between giving advice based on limited data or leaving the decisions to others.

In contrast to these three textbooks, the three-volume *Encyclopedia of Environmental Biology* is designed not for students but for professional biologists and government officials at all levels. William Nierenberg's desire to compile the encyclopaedia arose from his involvement with the issues of acid rain and global warming. His experiences convinced him that "governments continue to spend vast sums of money on research on both of these problems... but the smallest fraction is spent in the area of greatest concern, the biological environment". Presumably his encyclopaedia aims at improving this state of affairs by enlightening those public servants who consult its pages.

According to the editor's notes, the entries — from acid rain to zoological parks — were chosen to cover environmental topics of public and research

interest, including environmental issues important to lawyers. The needs of lawyers account for the inclusion of material on the spotted owl *Strix occidentalis caurina* and the Mount Graham red squirrel *Tamiasciurus hudsonicus grahamensis* (both subjects of much legal dispute in the United States) but the rationale behind other choices, such as the biology of mammalian infanticide, is obscure.

In any case, the authors and editor have succeeded in making the material accessible to the nonspecialist reader. The writing is generally clear and to the point and each article is followed by a glossary to assist those unfamiliar with the topic as well as by a short bibliography. The encyclopaedia should be helpful to government officials and lawyers, if they learn of its existence. But despite its illustrious advisory board, I doubt if most conservation biologists will find it essential because they are familiar with other

sources for most of the topics discussed.

In sum, *A Primer of Conservation Biology* and *Principles of Conservation Biology* would serve our hypothetical student well, while *Conservation Biology in Theory and Practice* and the *Encyclopedia of Environmental Biology* would be less useful. Although the authors of all four works are clearly committed to conservation, Meffe and Carroll are putting their money where their mouths are. Recognizing that conservation requires money as well as scientific insight, they and their publisher have promised to donate a third of the royalties from their text to major conservation organizations. They challenge other authors and publishers of works on ecology and conservation biology to do the same. I hope their challenge is met. □

Katherine Ralls is at the National Zoological Park, Smithsonian Institution, Washington DC 20008, USA.

Seeing the light of day (at last)

Peter Knight

Optical Coherence and Quantum Optics.

By Leonard Mandel and Emil Wolf. Cambridge University Press: 1995. Pp. 1,166. £30, \$49.95.

BEFORE the invention of the laser, one could rightly describe optics as being concerned with the manipulation of electromagnetic noise: natural light derived from the independent emission of many uncorrelated photons from atomic sources excited at random. The laser changed all that. For the first time, stimulated emission of radiation gave us a light source with coherence: that is, a predictable and controlled behaviour.

Leonard Mandel and Emil Wolf were the pioneers in the study of the statistical properties of light, starting well before the advent of the laser with their examination of the ground-breaking intensity correlation experiment of Hanbury Brown and Twiss on photon bunching. This, their long-awaited monograph on quantum optics, is the culmination of decades of study of the fluctuating nature of light and especially of its quantum properties. (Long-awaited indeed: the book was promised in the preface of the famous text of Max Born and Emil Wolf on classical optics in the 1950s, and I well remember seeing the initial stages of this book as a postdoc in Rochester more than 20 years ago.) Wolf was the originator of many of our fundamental ideas of coherence, and with Mandel was the founder of the experimental study of nonclassical light and the first to generate laboratory sources of light of totally

quantum origin such as anti-bunched light and the correlated twin beams of light from parametric amplifiers. Well, the wait was well worthwhile: they have delivered a masterly treatment of the subject, one that will delight professionals and students alike.

The book begins with a detailed account of the stochastic nature of light fluctuations and the coherence properties of light. The opening chapters describe how the mathematical theory of random processes, and especially Fokker-Planck and Langevin equations, can be applied to light fields; temporal and spatial coherence are described in terms of correlation functions. These are then related to spectra and applied to propagating light in radiometry and to light scattered from random media. Fundamental concepts of field quantization are treated next, with due weight given to the basic building blocks of optical fields; the number states with precise numbers of photons, and the coherent states that most closely resemble classical fields of well-defined amplitude and phase. Until recently, fundamental quantum properties of light were masked by environmental fluctuations so that for most purposes a careful semiclassical theory of light agreed in entirety with quantum treatments until new developments in the 1970s required fully quantized theories; underlying this agreement is the "optical equivalence principle" that relates semiclassical and quantum treatments.

The central part of the book contains a careful analysis of the quantum theory of light, the nature of photons and their

generation from atomic sources, ranging from single atoms in resonance fluorescence to laser oscillators and amplifiers. As one would expect from these authors, the treatment is a seamless fusion of high-level theoretical and experimental ideas, with a particularly good treatment of quantum noise in open optical systems and of the quantum theory of the laser. Finally, the last sections of the book deal with entirely nonclassical features of the kind of light fields generated recently, such as the phase-dependent noise of squeezed light, that justify the label 'quantum' in quantum optics. The text concludes with a discussion of the highly correlated two-photon states produced by parametric amplifiers and used to such effect by Mandel and others to demonstrate the optical consequences of nonlocality in quantum mechanics described by Bell's inequalities.

The level struck by the authors is right for students as well as professionals, and combines rigour with clarity and thoroughness. The authors concentrate on fundamental concepts and theoretical understanding, but throughout the book they bring out the links with experiments and the real world. The book is beautifully produced. When most serious monographs seem to be priced at extraordinary levels, it is a real delight to see the modest price of this text, which is bound to be a classic and moreover one that students and young researchers will be able to afford. □

Peter Knight is at the Blackett Laboratory, Imperial College of Science, Technology and Medicine, London SW7 2BZ, UK.

New in paperback

The Fourth Discontinuity: The Co-Evolution of Humans and Machines

by Bruce Mazlish. Yale University Press, £9.50. "Mazlish's book is most useful as a work of reference: a brief encyclopaedia of what various historical figures have already written on the subject". Harry Collins in *Nature* **367**, 606 (1994).

Essential Reproduction by Martin H. Johnson and Barry J. Everitt (4th edition). Blackwell Science, £19.50.

Transition Metal Oxides: An Introduction to their Electronic Structure and Properties by P. A. Cox. Oxford University Press, \$39.95.

A Treatise on the Theory of Bessel Functions by G. N. Watson, Cambridge University Press, £19.95, \$29.95. "... a monument of erudition... a rigorous mathematical treatment of all types of Bessel function", wrote L. M. Milne-Thompson in *Nature*.

Thirteen: The Apollo Flight That Failed by Henry S. F. Cooper. John Hopkins University Press, \$13, £11. First published in 1972.

The first 85 million years of avian evolution

Luis M. Chiappe

More than half of the evolutionary history of birds is played out in the Mesozoic. A recent burst of fossil discoveries has documented a tremendous diversity of early avians. Clarification of the Phylogenetic structure of this diversity has provided clues for a better understanding of the evolution of functional, developmental and physiological characteristics of modern birds. Yet their long Mesozoic history is only beginning to be deciphered.

THE spectacular specimens of the Late Jurassic *Archaeopteryx*, and to a lesser extent the more derived and much younger hesperornithiforms and ichthyornithiforms, have formed the core of our understanding of avian origins and early history for more than a century¹. Today, numerous new findings of Mesozoic birds—the number of taxa of which have doubled in the past five years—have shed light on the large chronological and phylogenetic gap between *Archaeopteryx* on the one hand and hesperornithiforms and ichthyornithiforms on the other, and research on early avian evolution has flourished with new ideas.

The theropodan origin of birds, namely the hypothesis that birds are most closely related to bipedal, carnivorous dinosaurs, is strongly supported by the known evidence^{2–4}. Yet, a few skeptics still regard crocodylomorphs^{5,6}, basal archosaurs⁷ or even a variety of non-archosaurian diapsids^{7,8} (such as *Megalanosaurus*, *Longisquama*) as the closest avian relatives. These arguments have consistently relied on excessive weights for a few similarities^{5–7} (such as inner ear, periotic pneumaticity)—found to be of broader distribution among archosaurs (including dinosaurs)—or on pointing at the presence of ‘bird-like’ structures in one or other taxon⁸ (such as feather-like scales in *Longisquama*, beak-like snout in *Megalanosaurus*). None of these interpretations, however, has furnished more than an *a priori* argument of convergence to explain the enormous body of morphological evidence supporting the notion that birds are nested within theropod dinosaurs.

A wealth of new Mesozoic birds has provided data for elucidating the series of branching events and structural transformations from non-avian dinosaurs to their extant descendants (Fig. 1), reinforcing the idea that modern birds are short-tailed, feathered dinosaurs. Coupled with a surge in physiological, functional and developmental studies on non-avian theropods and birds, these phylogenetic developments have illuminated new frontiers in early avian evolution. Here I will review these advances and discuss their implications for understanding the evolution of the biology of modern avians.

Archaeopteryx and the earliest bird

The known history of birds starts in the Late Jurassic (Fig. 1). Although presumably this history predates the end of this period, attempts to unravel its earlier phases have failed to provide reliable evidence. In recent years, two sets of fossils have been suggested to substantiate the record of birds before the Late Jurassic: *Protoavis* from the Late Triassic of Texas⁹ and an ensemble of traces from the Triassic–Jurassic of Africa^{10,11} and North America¹¹.

Chatterjee⁹ not only considered *Protoavis* a bird but also regarded it as much more closely related to modern, neornithine birds than is *Archaeopteryx*. Except for a few elements, the available material of *Protoavis* is extremely fragmentary. Chatterjee's interpretations of certain bones (such as furcula, sternum) are questionable, and even the association of elements into specimens and then into a single taxon seems difficult to support. Despite detailed reconstructions of the skull⁹, only portions of the braincase, quadrate and orbital roof are reasonably pre-

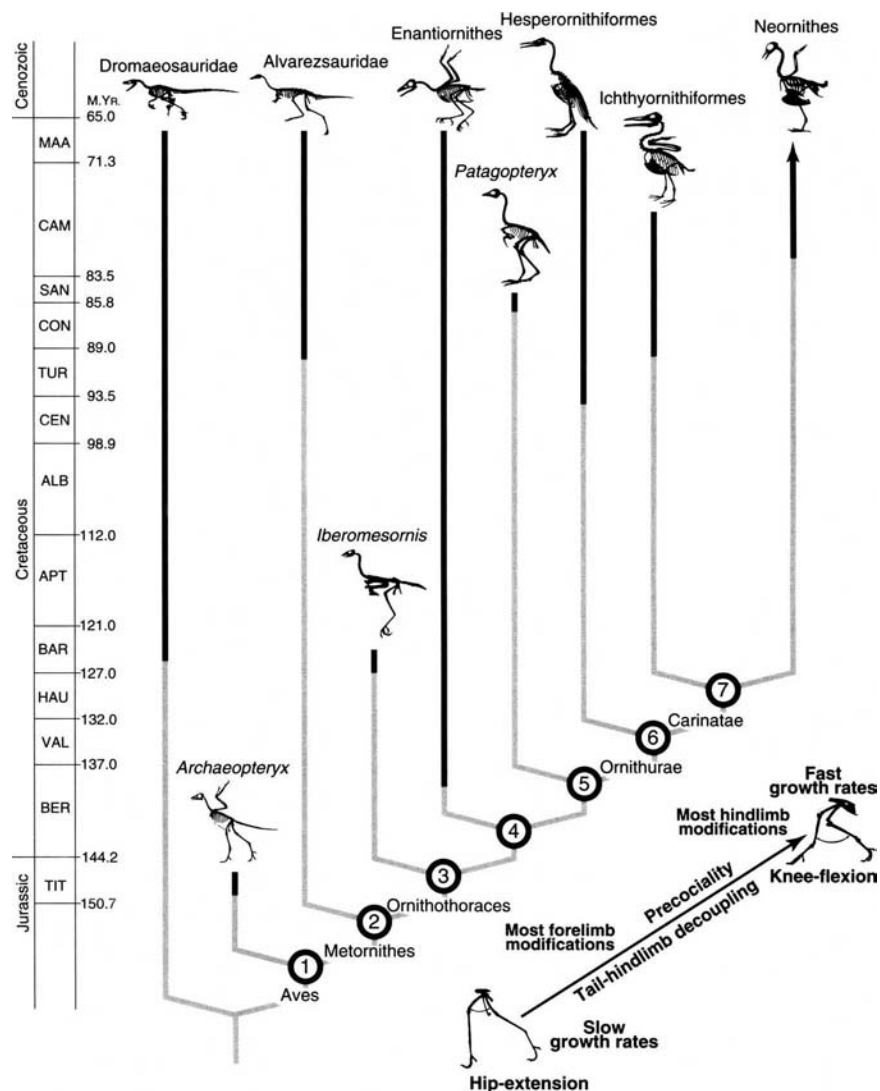
served. Nevertheless, the anterior cervical vertebrae approach a heterocoelic condition and dorsal vertebrae have large vertebral canals, both avian synapomorphies within theropods¹². Despite this, I concur with others¹³ that until better specimens are recovered and support for association of these into a single taxon provided, *Protoavis* should not be considered relevant to avian evolution.

In the early 1970s, Ellenberger¹⁰ regarded as avian a variety of Late Triassic–Early Jurassic footprints and traces from South Africa. More recently, Lockley *et al.*¹¹ reported on other Early Jurassic bird-like footprints from northern Africa and North America, suggesting again that these traces may represent evidence of birds before the Late Jurassic. Although these footprints deserve serious consideration, it is well known that the association of a footprint with a particular tracemaker always involves some degree of uncertainty, because several factors influence its shape and therefore its taxonomic identification¹⁴. Given the lack of reliable osteological evidence before the Late Jurassic, caution is advised in identifying these footprints as avian.

Thus, the presence of birds before the end of the Jurassic is far from documented. Yet, the singular position of *Archaeopteryx* as the oldest bird has been challenged by recent reports of alleged Late Jurassic avians. These new specimens have only been preliminarily studied or reported informally. A feathered specimen from North Korea was dubbed the ‘North Korean *Archaeopteryx*’¹⁵. Differences in its wing proportions suggest that this specimen is not *Archaeopteryx*, and its stratigraphic provenance has yet to be announced. *Confuciusornis*, from the Yixian Formation of northeastern China, is another purported Late Jurassic bird^{16,17}. The holotype is known from a skull associated with a wing. Other specimens from the same locality have been referred to this taxon^{16,17}, but lack of overlapping elements between these and the holotype make this assignment conjectural. Furthermore, no consensus exists on the age of the Yixian Formation, and palynological data support a younger, Early Cretaceous, age¹⁸. Despite the importance of these birds, further chronostratigraphical studies are necessary before placing confidence in their alleged age.

The spectacular specimens of *Archaeopteryx*¹ are still the most enlightening testimony of Late Jurassic birds. However, aspects of its biology remain controversial. Important anatomical complexes such as the braincase, orbit and palate are uncertain, and even whether these specimens belong to one or more closely related species is still controversial. In recent years, the burden of the debate has been concentrated on its aerodynamic ability and mode of life. Some opinions regard *Archaeopteryx* as primarily a terrestrial animal, with little or no capability for flight and a morphology and ecology similar to non-avian theropods such as *Deinonychus*^{3,19,20}. Others see it as an arboreal bird, with a design for flight approaching the condition of modern birds^{6,21,22}. Despite this broad range of opinions, most authors would agree that *Archaeopteryx* was capable of some sort of flight^{1,23–26}. The popular notion of an essentially arboreal *Archaeopteryx*—so frequently used in adaptational models of

FIG. 1 Phylogenetic relationships of the best-known taxa of mesozoic birds (data from Chiappe and Calvo¹²; Chiappe *et al.*⁶⁰; 146 steps, consistency index = 0.75, retention index = 0.81). Solid bars indicate the actual record of each lineage. Phylogenetic calibration of this record implies that most lineages must have had their origin at the beginning of the Cretaceous or even earlier. For example, Turonian ichthyornithiforms imply that, even though reliable evidence of neornithines is Campanian, we should expect to find earlier members of this clade in the Turonian. Indeed, the Early Cretaceous ornithurines *Enaliornis* and *Ambiortus* (here not included because their relationships are not clear) indicate that carinate birds, as ornithurines, must have been differentiated at least in the Early Cretaceous. Nodes 1 to 7 are diagnosed by the following unequivocal synapomorphies among others, NODE 1 (Aves = Aviale *sensu* Gauthier³): fewer than 26 caudals, caudals with short prezygapophyses, teeth with unserrated crowns and crown-base constrictions, completely reverted hallux; NODE 2 (Metornithes²⁹, the common ancestor of *Mononykus* and Neornithes plus all its descendants): prominent ventral processes on cervico-dorsals, carinate and longitudinally rectangular sternum, carpometacarpus; NODE 3 (Ornithothoraces¹², *Iberomesornis*, Neornithes and all descendants of their common ancestor): pygostyle, strut-like coracoid, sharp caudal end of scapula, radial shaft/ulnar shaft ratio smaller than 0.7; NODE 4: synsacrum with more than eight vertebrae, heterocoelous cervicals, complete calcaneo-astragalar-tibial fusion, distal tarsals completely fused to metatarsals; NODE 5: quadrate with three distal condyles forming a triangle, caudal prezygapophyses absent or extremely reduced, absence of pubic foot, tarsometatarsal vascular distal foramen; NODE 6 (Ornithurae^{5,42,69}, all birds derived from the common ancestor of Hesperornithiformes and Neornithes): sharp and pointed quadrate orbital process, fewer than 11 dorsals, procoracoid process, small acetabulum, pubis parallel to ischium and ilium, femur with prominent patellar groove; NODE 7 (Carinatae^{5,42}, Ichthyornithiformes, Neornithes, and all descendants from their common ancestor): globe-shaped, cranio-caudally convex humeral head. Tit, Tithonian; Ber, Berriasian; Val, Valanginian; Hau, Hauterivian; Bar, Barremian; Apt, Aptian; Alb, Albian; Cen, Cenomanian; Tur, Turonian; Con, Coniacian; San, Santonian; Cam, Campanian; Mas, Maastrichtian (boundaries between stages follow Gradstein *et al.*⁸⁵).



the origin of flight^{21,22}—has encountered resistance from several detailed morphological studies^{19,24,27}. Moreover, this idea has consistently relied on a palaeoenvironmental misconception, for the Solnhofen lagoons were not surrounded by forests but by sparse small plants which at the most reached 3 m in height²⁸. Phylogenetic reconstruction does not provide any direct evidence about the specific aerodynamic ability or mode of life of *Archaeopteryx*. Nevertheless, it strongly supports that *Archaeopteryx* was morphologically very different from its living relatives (Fig. 1), and that the ancestral mode of life for birds was predominantly terrestrial as it must have been for their closest theropod outgroups and other basal avians (such as *Mononykus*²⁹).

The Cretaceous diversity

Cretaceous birds have often been regarded as early members of extant lineages^{30–32}. Only recently have these fossils been interpreted as remotely connected to modern groups and the enormous Cretaceous diversity fully appreciated.

The most informative specimens of Early Cretaceous birds come from Spain, China and Mongolia (Fig. 2). Although several of these Early Cretaceous birds can be placed within the Enantiornithes—a diverse, worldwide group of Cretaceous volant birds—a number of them represent lineages so far known only by a single species.

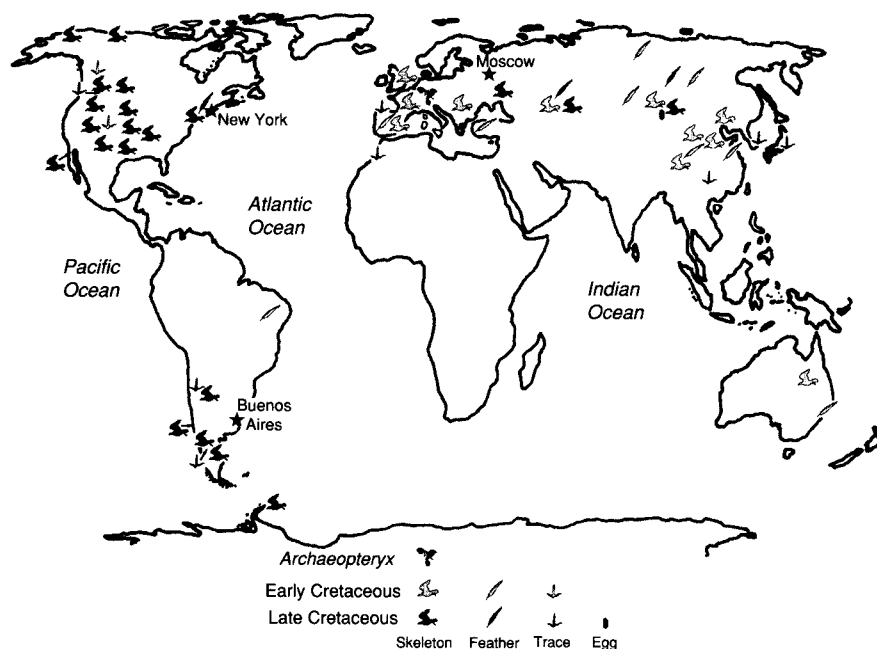
Aptian; Alb, Albian; Cen, Cenomanian; Tur, Turonian; Con, Coniacian; San, Santonian; Cam, Campanian; Mas, Maastrichtian (boundaries between stages follow Gradstein *et al.*⁸⁵).

The lowermost Cretaceous limestones of Montsec (northern Spain) have yielded the remains of the finch-sized *Noguerornis*³³. Although the single known specimen is fragmentary, *Noguerornis* documents important structural novelties correlated with enhanced flying ability, such as interlocking of manual elements into a rigid structure, elongation of the distal portions of the wing, and enlargement of the wing surface, only a few million years after *Archaeopteryx*. Assessment of its phylogenetic relationships is complicated because it is poorly preserved. Nevertheless, retention of several plesiomorphies (such as ischiadic symphysis, subequal widths of radius and ulna) suggests a sister-taxon relationship to all Ornithothoraces¹².

More complete specimens of other taxa were found in slightly younger rocks from central Spain. The Barremian limestones from Las Hoyas have yielded the basal *Iberomesornis*³⁴ and the enantiornithine *Concornis*³⁵ (Fig. 3). Sparrow-sized *Iberomesornis* is the most basal Ornithothoraces (Fig. 1). Like *Noguerornis*, and in contrast to *Archaeopteryx*, *Iberomesornis* shows characters (such as strut-like coracoid, ulna longer than humerus) suggesting an improved flying capability³⁴. It also shows a fully specialized perching foot, indicating that arboreality and perching capacity arose very early in avian history.

Remains of Early Cretaceous birds have been found in several Chinese localities (Fig. 2). Most informative specimens come

FIG. 2 Geographical distribution of Mesozoic birds. Note the relatively limited record of the southern continents, in particular Africa. *Archaeopteryx* is here regarded as the only indisputable evidence of pre-Cretaceous birds (see text for further discussion).



from the lacustrine deposits of the Valanginian Jiufotang Formation at Liaoning Province^{36,38}, including the sparrow-sized *Sinornis*³⁶ and *Cathayornis*³⁷, which are unquestionably enantiornithines. Other less complete specimens, such as the larger *Chaoyangia*³⁸, are more difficult to place in a particular clade. Despite the retention of plesiomorphies such as unfused pelvic elements and a pubic symphysis, *Chaoyangia* stands out among early avians in possessing large, ossified uncinat processes otherwise unknown in the ribs of any non-ornithurine bird¹² (Fig. 1).

Known Early Cretaceous birds are relatively small and most of them have perching, arboreal specializations (such as *Iberomesornis*, *Sinornis*, *Cathayornis*, *Concornis*). At the end of the Early Cretaceous and during the Late Cretaceous, however, the

known fossil specimens range widely in size and life style, from flightless cursors and foot-propelled divers to waders and tree-dwellers. Several of these lineages belong to the Enantiornithes, but many of them are members of the Ornithurae (Fig. 1).

Two remarkable, flightless, ground-dwelling lineages have their first records in the Late Cretaceous: the central Asian, stout-forelimbed *Mononykus*²⁹ and its South American relatives (see below), and the Patagonian *Patagopteryx*³⁹. Hen-sized *Patagopteryx* was originally regarded as a primitive ratite³⁹, though the absence of many derived features of ornithurine birds (Fig. 1) indicates that it is neither a ratite nor a member of the Ornithurae; *Patagopteryx* is the sister-group of all ornithurine birds^{12,40}. *Patagopteryx* acquired flightlessness independent from all other birds, and stands out among birds in having a fused

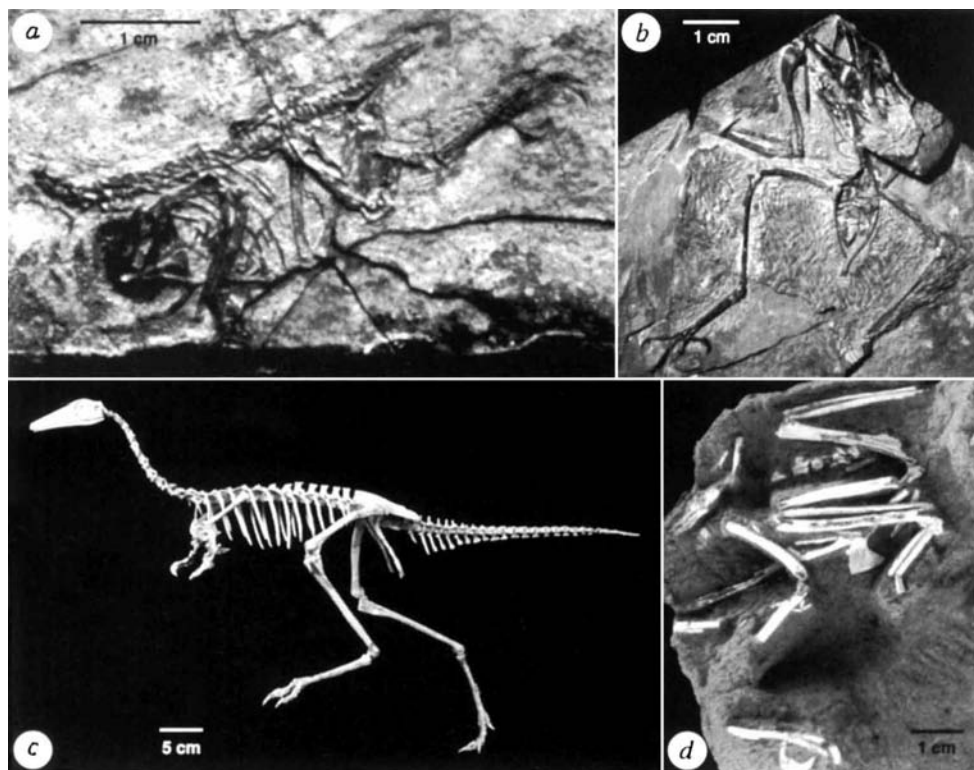


FIG. 3 a, b, *Iberomesornis* and the enantiornithine *Concornis*, respectively, from the Early Cretaceous (Barremian) of Las Hoyas, central Spain. c, Skeletal mounting of *Mononykus* from the Late Cretaceous (Campanian-Maastrichtian) of southern Mongolia. d, Enantiornithine *Neuquenornis* from the Late Cretaceous (Coniacian-Santonian) of northwestern Patagonia, Argentina.

quadrate-ptyergoid complex, which suggests a unique kind (if any) of cranial kinesis.

Reconstructing the early history of the Ornithurae is hampered by the fragmentary nature of their putative earliest members. The Early Cretaceous *Ambiortus*⁴¹ from Mongolia is probably the oldest ornithurine. Nevertheless, its position within the Ornithurae is not yet clear^{25,42}; the presence of amphicoelus vertebrae might suggest a close relationship with ichthyornithiforms. Sereno and Rao³⁶, however, have placed *Ambiortus* outside Ornithurae, but they have not presented evidence (character distribution) supporting their claim.

The best known Mesozoic ornithurines are the hesperornithiforms: toothed, foot-propelled, flightless divers known primarily from marine Late Cretaceous rocks of North American Western Interior^{5,25,43}. In recent years, these birds have also been recorded in rocks of the Late Cretaceous Turgai strait⁴⁴, between eastern Europe and western Asia. Remains of these birds have been found in continental⁴⁵ and estuarine⁴⁶ deposits, indicating that they also inhabited non-marine environments. Whether *Enaliornis*⁴⁷ from the Early Cretaceous of the UK is a primitive member of this group is not yet clear, for despite similarities in pedal morphology⁵, the braincase appears to resemble modern birds more than *Hesperornis*⁴⁷. Hesperornithiforms have often been regarded as primitive grebes and loons^{30,31,48}, or as aquatic palaeognaths³². Recent studies, however, have consistently regarded them as basal ornithurines^{5,12,42}, and their flightlessness interpreted as a secondary specialization (Fig. 1). Elzanowski⁴⁹, however, has suggested a closer connection to neognaths, namely as the sister-group of all modern birds except palaeognaths (ratites and tinamous). Nevertheless, the latter hypothesis requires explanation of additional homoplasies, such as the independent loss of teeth in both palaeognaths and neognaths.

Also from the Late Cretaceous of the North American Western interior^{5,25,43} are the flying, toothed ichthyornithiforms. Claims reporting the occurrence of these birds in the Late Cretaceous of Asia⁴⁴ have been recently refuted⁴⁵. Though known since the last century, few papers have dealt with them after Marsh's 1880 monograph⁴³. Earlier authors saw them as primitive charadriiforms, although most recent phylogenetic studies have placed them outside modern birds^{5,42} (Fig. 1).

There is compelling evidence documenting the presence of several lineages of modern birds at the end of the Cretaceous. Taxa closely related to extant anseriforms⁵⁰, gaviiforms⁵¹, charadriiforms^{52,53} and procellariiforms^{44,45,53} have been found in Campanian–Maastrichtian rocks, and phylogenetic inferences predict that these and other lineages (such as palaeognaths) must have differentiated even earlier (Fig. 4). The documented earlier history of Neornithes, however, is obscured again by the fragmentary preservation of their putative oldest representatives. Early Cretaceous *Gansus*⁵⁴ from China and *Paleocursornis*⁵⁵ from Romania have been related to modern charadriiforms and ratites, respectively, but the evidence presented so far is inconclusive.

The *Mononykus* problem

Perhaps the most startling recent finding is the flightless, stout-forelimbed, turkey-sized *Mononykus* from the Late Cretaceous of central Asia^{29,56} (Fig. 3). Close relatives of *Mononykus* lived in the Late Cretaceous of Argentina⁵⁷ and probably western North America⁵⁸. Novas⁵⁷ has reinterpreted the Patagonian *Alvarezsaurus*—formerly thought to be a non-avian theropod⁵⁹—as a bird, and placed *Mononykus* and a new Argentine taxon within the monophyletic Alvarezsauridae (Fig. 1).

Placement of these bizarre creatures within Aves has created a great deal of controversy. Nevertheless, aside from its grotesque forelimbs, *Mononykus* and its relatives exhibit several avian synapomorphies and other derived characters only known for birds more derived than *Archaeopteryx*^{29,56,60} (Fig. 1). Perhaps its most manifest avian characteristic has been documented by a recently found complete skull, in which the orbit is connected to the

infratemporal fenestra, a condition exclusively known for birds among archosaurs. The placement of *Mononykus* outside Aves requires the less parsimonious conclusion that all these characters evolved independently in these two taxa. This fact has not avoided criticism in which *Mononykus* is *a priori* disregarded as a bird^{61–64}. This criticism seems to stem from the conjecture that *Mononykus* does not fit the 'stereotype' of a basal bird⁶⁵. For example, Wellnhofer⁶¹ finds it 'very difficult to imagine how a primitive bird wing, such as that of *Archaeopteryx*, could have evolved into a forelimb like that of *Mononykus*', and Feduccia⁶² claims that 'the keeled breastbone doesn't resemble that of birds . . . and it is probably associated with a digging function, not any birdlike actions'. However, assumptions about evolutionary processes or adaptational scenarios, such as these, are misleading when identifying historical relationships. Phylogenetic reconstruction should be based exclusively on the hierarchical distribution of homologies among taxa, and a given phylogenetic hypothesis can only be rejected by providing an alternative hypothesis for which supporting evidence outweighs that of the original hypothesis. Clearly, such a procedure has not been implemented by those that reject out-of-hand the avian relationship of *Mononykus*⁶⁵.

The Enantiornithes: an unexpected radiation

Distributed worldwide, ranging from the Valanginian to the Maastrichtian, and known by over a dozen distinct species, the Enantiornithes document a major cladogenetic event of Mesozoic volant birds (Fig. 1). Quite surprisingly, this clade remained unrecognized until 1981⁶⁶. Enantiornithine fossils had been found earlier—even a century ago—but they remained misidentified as non-avian theropods (such as *Ornithomimus*) or were included within modern groups (such as, *Gobiapteryx*⁶⁷).

Except for a few occurrences in near-shore deposits, most enantiornithines are known from non-marine rocks. Most likely, this bias in the fossil record reflects their preponderance in Cretaceous terrestrial avifaunas. Despite the fact that all enantiornithines show morphological features suggesting an enhanced flying capacity¹², they exhibit a broad morphological range. For example, although early forms such as *Cathayornis* and *Sinornis* were tiny and toothed, late Mesozoic taxa such as *Enantiornis*⁶⁶ reached a wing-span of over 1 m, and some like *Gobiapteryx*⁶⁷ lacked teeth. Interpretation of *Sinornis* as an enantiornithine modifies Sereno and Rao's hypothesis regarding this taxon as the most basal bird apart from *Archaeopteryx*³⁶. Although these authors did not provide support for their hypothesis, *Sinornis* shows several derived characters that connect it to the Enantiornithes (such as reduced metatarsal IV, dorsal process on ischium, parapophyses of dorsal vertebrae centrally located). One of the most striking structural characteristics of the Enantiornithes has been revealed by histological studies of their bones, which have shown a unique pattern when compared with those of other birds⁶⁸.

The relationship of the Enantiornithes to other Mesozoic avian lineages has been the subject of intense debate^{40,69}. This clade has been regarded variously as closely related to *Archaeopteryx*^{5,6}, included within the Ornithurae⁴², or placed in an intermediate position between *Archaeopteryx* and ornithurine birds^{12,40,66,69}. The latter hypothesis is clearly the one that best summarizes the available evidence (Fig. 1). The sister-group relationship between *Archaeopteryx* and Enantiornithes has been consistently argued by Martin^{5,6}, who has envisioned a basal avian dichotomy between the 'Sauriurae' (Enantiornithes and *Archaeopteryx*) and its alleged sister-taxon, the Ornithurae. The 'Sauriurae', however, is a paraphyletic group supported by characters that are erroneous or, at best, plesiomorphic⁴⁰. This unjustified dichotomy turns the otherwise monophyletic Enantiornithes into a wastebasket for everything that is neither *Archaeopteryx* nor Ornithurae. This arrangement has become even more complex because of indiscriminate use of paraphyletic taxa as practiced by Feduccia⁷⁰ who, though endorsing the 'Sauri-

urae', represents the Enantiornithes as more closely related to Neornithes than to *Archaeopteryx*, and regards the Ornithurae as a terminal taxon not including modern, neornithine birds.

Growth, developmental modes and physiology

Comparative studies on bone microstructure have been extensively used to infer growth rates, and in some instances, physiological aspects of extinct vertebrates. The bone histology of Mesozoic birds is known only for a few taxa. In *Hesperornis*, the bone tissue is like that of its extant counterparts⁷¹, with uninterrupted, fast-deposited fibro-lamellar bone. This type of bone is consistent with rapid growth; modern birds grow rapidly, reaching adult size within the first year⁷². A different pattern, however, has been found in Enantiornithes and *Patagopteryx*. Microstructural studies of these birds have documented punctuations interrupting bone deposition (lines of arrested growth or LAGs), suggesting cyclical pauses during post-natal growth^{68,73}. Such a pattern contrasts with the uninterrupted growth pattern of extant birds. Furthermore, although *Patagopteryx* shares the modern condition of a highly vascularized, fibro-lamellar bone, the enantiornithines exhibit a lamellar, slowly deposited bone tissue, with virtually no vascularization^{68,73}, a pattern remarkably different from that of modern birds and their non-avian theropod relatives.

Most probably, LAGs of Enantiornithes and *Patagopteryx* formed annually, as occurs in extant non-avian reptiles⁷⁴. This pattern of discontinuous growth, and the slowly deposited bone tissue of Enantiornithes, suggest that these birds had slower rates of growth than their modern relatives. Four and five LAGs observed in two enantiornithine specimens indicate that these individuals had at least four and five years, respectively, of post-hatching growth^{68,73}. Moreover, the fact that six specimens of *Archaeopteryx* were found to lie on a linear size trajectory⁷⁵ also hints at slow rates of post-natal growth for basal birds, because it would be very unlikely that all these *Archaeopteryx* individuals died differing by only a few months of age.

The inferred growth rates for these birds shed light on their developmental modes. Hatchlings of extant birds form a broad spectrum of appearance and conduct, ranging from downy, precocial species capable of independent locomotion to the usually naked and blind altricial taxa that are incapable of locomotion^{72,76}. Precociality has been traditionally regarded as primitive for it is typical of basal neornithines such as palaeognaths and galliforms⁷². Evidence of precociality in basal Mesozoic birds was suggested by the discovery of embryos of the enantiornithine *Gobipteryx* showing a remarkable degree of ossification⁷⁷, a condition also found in non-avian theropod embryos⁷⁸. The slow growth rates inferred for basal birds are consistent with this hypothesis, for living precocial taxa have slower rates of growth than do altricial birds⁷².

Observed interruptions during bone deposition of Enantiornithes and *Patagopteryx* also hint at other physiological differences with respect to modern birds. Growth, as is the case for other activity levels, is strongly correlated with body temperature⁷⁹. Hence, inference of cyclical growth in basal birds suggests they may not have been endothermic homeotherms^{68,73}. The fact that these basal birds may not have been endotherms does not necessarily imply that they were ectotherms, for they conceivably may have had intermediate levels of activity as it has been suggested for non-avian dinosaurs⁷⁴. If this is correct, however, characterizations of *Archaeopteryx* and the enantiornithine *Sinornis* as endotherms should be re-evaluated^{68,73,80}.

Interpretation of these primitive birds as non-endothermic conflicts with ideas envisioning the origin of feathers in conjunction with endothermic metabolism. For example, Ostrom¹⁹ maintained that endothermy 'must have been achieved before there was any extensive feather covering' and that it 'must have preceded flight', but the Enantiornithes were covered by feathers and its anatomy suggests an enhanced flying ability¹². Furthermore, not all feathered creatures are endothermic (downy hatch-

lings are unable to generate internal heat⁷⁶) nor does any kind of flight appear to be necessarily related to endothermy^{80,81}. Although some of these aspects may be conjectural, the histological differences documented for early birds suggest that important physiological changes may have occurred late in avian history. The notion that classic, modern endothermy may have evolved after the development of feathers and an enhanced flying ability should be seriously considered^{68,73,80}.

Terrestrial locomotion and flight apparatus

Birds are unique among extant vertebrates in that fore- and hindlimbs are part of two independent locomotory systems. Functional limb decoupling certainly took place long before the origin of birds, for bipedalism is ancestral even for dinosaurs. The modern avian pattern of hindlimb kinematics, however, differs from the ancestral theropod one. As Gatesy⁸² has shown, in non-avian theropods—as in lizards and crocodiles—hip-extension was the primary mechanism of hindlimb kinematics, namely a pattern of extensive femoral retraction during each stride. In contrast, a knee-flexion mechanism is characteristic of modern birds, in which the stride is coupled by a large tibiotarsal displacement⁸². Most probably, reduction of the tail and caudal musculature throughout theropod evolution is the principal agent implicated in the development of the latter mechanism. Correlated with this transformation are the functional decoupling of the tail from the primitive pattern of hindlimb kinematics and its linkage with the flight apparatus, and the forward migration of the centre of gravity and development of a modern avian stance⁸². Interestingly, these transformations were not fully achieved until late in avian evolution. The relatively long tails of *Archaeopteryx* and *Mononykus* suggest that these basal birds retained a pattern of hindlimb kinematics similar to that of non-avian theropods. Furthermore, although the pygostyle and relatively short tails of early ornithothoracines (such as *Iberomesornis*) indicate tail decoupling, retention of several primitive pelvic characters suggests that in these birds the transformation was not yet completed. For example, pubic and ischiadic symphyses are present in *Concornis* and *Noguerornis*, respectively. *Patagopteryx* has an ample brevis fossa (origin of the *M. caudofemoralis brevis*) and a fairly long tail. In fact, it is only in the Ornithurae that the pelvis and hindlimb acquired a fully modern appearance. Thus, though the tail of several basal birds is poorly known and specific morphofunctional analyses are still pending, the modern pattern of hindlimb kinematics was probably not fully developed until the differentiation of the Ornithurae^{40,69}.

Conversely, a condition approaching a modern design of wing and shoulder girdle was acquired very early in avian history.

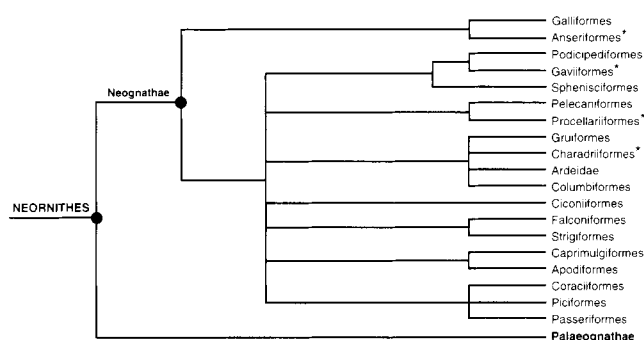


FIG. 4 Phylogenetic relationships among modern, neornithine birds (after Cracraft⁸⁶). Asterisked taxa are known to occur in Late Cretaceous deposits. Note that the phylogenetic pattern among these taxa suggests that several other lineages (such as palaeognaths) must have been differentiated before the end of the Cretaceous. Neornithes includes the common ancestor of living birds plus all its descendants; this clade is subdivided in Palaeognathae (ratites and tinamous) and Neognathae (all other extant birds).

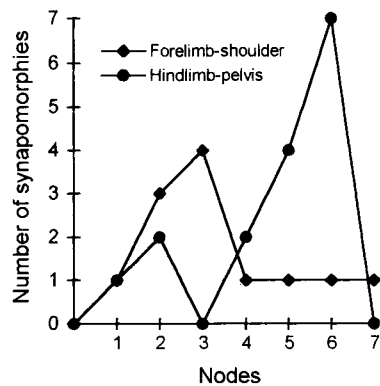


FIG. 5 Sequence of transformations of the forelimb and shoulder girdle (including sternum), and the hindlimb and pelvis during early avian phylogeny. The chart shows the number of synapomorphies (only unequivocally optimized) for each node of the cladogram in Fig. 1. The fact that most synapomorphies of the forelimb and shoulder girdle precede those of the hindlimb and pelvis suggests that during the early history of birds, the structures correlated to an enhanced flying ability were acquired earlier than those correlated to a modern kind of terrestrial locomotion⁴⁰. Character data derived from Chiappe et al.⁶⁰.

Although the aerodynamic capabilities of *Archaeopteryx* are still a matter of controversy, the anatomy of the slightly younger *Noguerornis* and *Iberomesornis* documents an enhanced flying ability. Thus, the evolution of the modern flight apparatus appears to have preceded the modern mechanism of hindlimb kinematics, a conclusion supported by the fact that the sequence of structural modifications leading to the former appear earlier (that is, are hierarchically more inclusive)^{40,69} (Fig. 5).

Temporal branching and taxonomic dynamics

Even among the most complete sequences, the fossil record can hardly ever be taken at face value⁸³. Despite the large number of recent discoveries, large segments of the early history of birds are still missing (Figs 1 and 6). The available fossils, however, can provide a clearer account of temporal patterns if they are calibrated using a phylogenetic hypothesis⁸⁴. The minimal age for a lineage and its sister-taxon can be retrieved from their

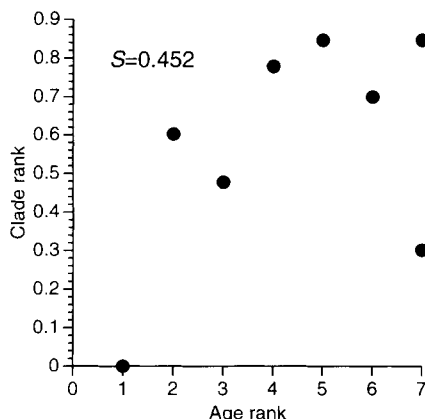


FIG. 6 Congruence between stratigraphical record and branching pattern for the cladogram of Fig. 1. Plots show age ranks (derived from the oldest record of terminal taxa) as a function of clade ranks (assigned to each branching event; for example, *Archaeopteryx* is 1, *Alvarezsauridae* is 2 and so on). Chart constructed by using the method described by Norell and Novacek⁸⁷. Clade ranks are rescaled from 0 to 1. The correlation is not even significant at $P < 0.05$. This low correlation indicates that despite the numerous new findings, the fossil record of basal birds is still incomplete and great portions of their evolutionary history are yet missing. S is the Spearman rank correlation coefficient.

oldest member, for sister-taxa must have a common origin. Calibrating the phylogeny of Mesozoic groups (Fig. 1) indicates that most, if not all, the Cretaceous lineages were present very early in this period⁴². Unfortunately this conclusion tells us nothing about the amount of time between the ancestor of all birds and those of each branching event. It shows, however, that recent claims proposing an explosive cladogenesis of modern birds at the base of the Tertiary⁷⁰ are inaccurate. Feduccia's new model⁷⁰ not only illustrates the misleading effects of using the fossil record at face value⁸⁴ but it also ignores evidence documenting several modern lineages in Late Cretaceous rocks^{50, 53} (Fig. 4). His interpretation of massive avian extinctions at the K-T boundary⁷⁰ is also paradoxical, for in this case the information available in the fossil record has been disregarded (Fig. 1). No non-neornithine Mesozoic lineages are known for the Tertiary, and most probably became extinct before the end of the Cretaceous. There is no positive evidence, however, that these extinctions occurred at the K-T boundary, for the last record of several lineages is before the Maastrichtian²⁵ (Fig. 1).

Prospects and conclusion

Birds flourished during the Mesozoic, acquiring specializations for a variety of different habitats. This was a period of extensive diversification, though one in which extinction was also a common phenomenon. The many new findings have provided additional support for the hypothesis that birds evolved from cursorial, bipedal carnivorous dinosaurs. Furthermore, these fossils have documented that several important morphological, functional and physiological systems characteristic of modern birds were acquired only late in avian history. Clearly, the common practice of extrapolating modern avian characteristics to these basal avian members or further down to non-avian dinosaurs, is at best a naive procedure.

Although the new fossil evidence has opened a realm of scientific inquiry, it has also exposed our relative ignorance about the first 85 million years of avian history. Further research is needed before a more comprehensive, finely tuned understanding of the interrelationships among the basal taxa can be developed and the importance of these fossils in clarifying the still controversial historical relationships of their extant relatives assessed. □

Luis M. Chiappe is in the Department of Vertebrate Paleontology and the Department of Ornithology, American Museum of Natural History, Central Park West at 79th Street, New York, New York 10024, USA.

- Hecht, M. K., Ostrom, J. H., Viohl, G. & Wellnhofer, P. (eds) *The Beginnings of Birds* (Jura Museum, Eichstätt, 1985).
- Ostrom, J. H. *Biol. J. Linn. Soc.* **8**, 91–182 (1976).
- Gauthier, J. *Mem. Calif. Acad. Sci.* **8**, 1–55 (1986).
- Witmer, L. in *Origin of the Higher Groups of Tetrapods* (eds Schultze, H.-D. & Trueb, L.) 427–466 (Comstock, Ithaca, 1991).
- Martin, L. D. in *Perspectives in Ornithology* (eds Bush, A. H. & Clark, G. A. Jr) 291–338 (Cambridge Univ. Press, 1983).
- Martin, L. D. in *Origin of the Higher Groups of Tetrapods* (eds Schultze, H.-D. & Trueb, L.) 485–540 (Comstock Publishing Associates, Ithaca, 1991).
- Tarsitano, S. in *Origin of the Higher Groups of Tetrapods* (eds Schultze, H.-D. & Trueb, L.) 541–576 (Comstock, Ithaca, 1991).
- Feduccia, A. & Wild, R. *Naturwissenschaften* **80**, 564–566 (1993).
- Chatterjee, S. *Phil. Trans. R. Soc. Lond.* **B332**, 277–346 (1991).
- Ellenberger, P. *Palaeovertebrata (Mem. Extraor.)* 1–141 (1974).
- Lockley, M. G., Yang, S. Y., Matsukawa, M., Fleming, F. & Lim, S. K. *Phil. Trans. R. Soc.* **B336**, 113–134 (1992).
- Chiappe, L. M. & Calvo, J. O. *J. Vert. Paleont.* **14**, 230–246 (1994).
- Ostrom, J. H. *Nature* **353**, 212 (1991).
- Padian, K. & Olsen, P. E. *J. Paleont.* **58**, 178–184 (1984).
- Korean Pictorial* **2**, 25 (1994) (in Japanese).
- Hou, L. in *Short Papers, IV Symp. Mes. Terrest. Ecosyst.* (eds Sun, A. & Wang, Y.) 193–201 (China Ocean, Beijing, 1995).
- Hou, L., Zhou, Z., Martin, L. D. & Feduccia, A. *Nature* **377**, 616–618 (1995).
- Li, W. & Liu, Z. *Cretaceous Res.* **15**, 333–365 (1994).
- Ostrom, J. H. *Q. Rev. Biol.* **49**, 27–47 (1974).
- Vazquez, R. *J. Res. Explor.* **8**, 387–388 (1992).
- Feduccia, A. & Tordoff, H. B. *Science* **203**, 1021–1022 (1979).
- Feduccia, A. *Science* **259**, 790–793 (1993).
- Olson, S. L. & Feduccia, A. *Nature* **278**, 247–248 (1979).
- Speakman, J. R. *Evolution* **47**, 336–340 (1993).
- Olson, S. L. in *Avian Biology Vol. 8* (eds Farner, D. S., King, J. & Parkes, K. C.) 79–238 (Academic, New York, 1985).
- Dodson, P. *Vert. Paleont.* **5**, 177–179 (1985).
- Peters, D. S. & Görgner, E. *Sci. Ser. nat. Hist. Mus. Los Angeles Co.* **36**, 29–37 (1992).

28. Barthel, K. W., Swinburne, N. H. M. & Morris, S. C. *Solnhofen: A Study in Mesozoic Paleontology* (Cambridge Univ. Press, 1990).
29. Perle, A., Norell, M. A., Chiappe, L. M. & Clark, J. M. *Nature* **362**, 623–626 (1993).
30. Howard, H. *Ibis* **92**, 1–21 (1950).
31. Brodkorb, P. in *Avian Biology* (eds Farner, D. S. & King, J. R.) 19–55 (Academic, New York, 1971).
32. Simpson, G. G. *Contrib. Sci. nat. Hist. Mus. Los Angeles Co.* **330**, 3–8 (1980).
33. Lacasa, A. *Estud. Geol.* **45**, 417–425 (1989).
34. Sanz, J. L. & Bonaparte, J. F. *Sci. Ser. nat. Hist. Mus. Los Angeles Co.* **36**, 39–49 (1992).
35. Sanz, J. L., Chiappe, L. M. & Buscalioni, A. D. *Am. Mus. Nov.* **3133**, 1–23 (1995).
36. Sereno, P. & Rao, C. *Science* **255**, 845–848 (1992).
37. Zhou, Z., Jin, F. & Zhang, J. *Chinese Sci. Bull.* **37**, 1365–1368 (1992).
38. Hou, L. & Zhang, J. *Vert. Palae.* **31**, 217–224 (1993) (in Chinese).
39. Alvarenga, H. M. F. & Bonaparte, J. F. *Sci. Ser. nat. Hist. Mus. Los Angeles Co.* **36**, 51–64 (1992).
40. Chiappe, L. M. *Courier Forschungsinstitut Senckenberg* **181**, 55–63 (1995).
41. Kurochkin, E. N. *Cretaceous Res.* **6**, 271–278 (1985).
42. Cracraft, J. *Paleobiology* **12**, 383–399 (1986).
43. Marsh, O. C. *United States Geological Exploration of the 40th Parallel* 1–201 (Government Printing Office, Washington, 1880).
44. Nesson, L. A. *Russ. J. Ornithol.* **1**, 7–50 (1993) (in Russian).
45. Kurochkin, E. N. in *Short Papers, IV Symp. Mes. Terrest. Ecosyst.* (eds Sun, A. & Wang, Y.) 203–208 (China Ocean, Beijing, 1995).
46. Fox, R. C. *Can. J. Earth Sci.* **11**, 1335–1338 (1974).
47. Elzanowski, A. & Galton, P. M. *J. Vert. Paleont.* **11**, 90–107 (1991).
48. Cracraft, J. *Syst. Zool.* **31**, 35–56 (1982).
49. Elzanowski, A. *Postilla* **207**, 1–20 (1991).
50. Noriega, J. I. & Tambussi, C. P. *Ameghiniana* **32**, 57–61 (1995).
51. Olson, S. L. *Vert. Paleontol.* **12**, 122–124 (1992).
52. Brodkorb, P. *Acta. XIII Congr. Int. Ornithol.* 55–70 (1963).
53. Olson, S. L. & Parris, D. Smith. *Contr. Paleobiol.* **63**, 1–22 (1987).
54. Hou, L. & Liu, Z. *Sci. Sin.* **27**, 1296–1302 (1984).
55. Kessler, E. & Jurcsák, T. *Trav. Mus. d'Hist. Nat. Grigore Antipa* **28**, 289–295 (1986).
56. Perle, A., Chiappe, L. M., Barsbold, R., Clark, J. M. & Norell, M. A. *Am. Mus. Nov.* **3105**, 1–29 (1994).
57. Novas, F. *Internatnl. symp. Gond. Dinos. Mem. Queens. Mus.* (Trelew, 1994).
58. Holtz, T. R. Jr. *J. Vert. Paleont.* **14**, 480–519 (1994).
59. Bonaparte, J. F. *Rev. Mus. Arg. Cien. Nat. "Bernardino Rivadavia" (Paleont.)* **4**, 17–123 (1991) (in Spanish).
60. Chiappe, L. M., Norell, M. A. & Clark, J. M. *Internatnl. symp. Gond. Dinos. Mem. Queens. Mus.* (Trelew, 1994).
61. Wellnhofer, P. *C. R. Acad. Sci. Paris* **319**, 299–308 (1994).
62. Feduccia, A. *Living Bird* **13**, 28–33 (1994).
63. Ostrom, J. H. in *Major Features in Vertebrate Evolution* (conv. Prothero, D. R. & Schoch, R. M.) *Short Course in Paleontology*, Vol. 7, 160–177 (1994).
64. Martin, L. D. & Rinaldi, C. *Maps Digest* **17**, 190–196 (1994).
65. Chiappe, L. M., Norell, M. A. & Clark, J. M. *C. R. Acad. Sci. Paris* **320**, 1031–1032 (1995).
66. Walker, C. A. *Nature* **292**, 51–53 (1981).
67. Elzanowski, A. *Palaeont. Pol.* **37**, 153–165 (1977).
68. Chinsamy, A., Chiappe, L. M. & Dodson, P. *Nature* **368**, 196–197 (1994).
69. Chiappe, L. M. *Alcheringa* **15**, 333–338 (1991).
70. Feduccia, A. *Science* **267**, 637–638 (1995).
71. Houde, P. *Auk* **103**, 125–129 (1987).
72. Starck, J. M. *Curr. Ornithol.* **10**, 275–366 (1993).
73. Chinsamy, A., Chiappe, L. M. & Dodson, P. *Paleobiology* **21** (in the press).
74. Chinsamy, A. & Dodson, P. *Am. Sci.* **83**, 174–180 (1995).
75. Houck, M. A., Gauthier, J. A. & Strauss, R. E. *Science* **247**, 195–198 (1990).
76. Campbell, B. & Lack, E. (eds) *A Dictionary of Birds* (Buteo, Vermillion, 1985).
77. Elzanowski, A. *Acta XVIII Congr. Int. Ornithol.* **1**, 178–183 (1985).
78. Norell, M. A. et al. *Science* **266**, 779–782 (1994).
79. Prosser, C. L. (ed.) *Comparative Animal Physiology* (Saunders, New York, 1973).
80. Randolph, S. E. *Zool. J. Linn. Soc.* **112**, 389–397 (1994).
81. Ruben, J. *Evolution* **45**, 1–17 (1991).
82. Gatesy, S. M. in *Functional Morphology in Vertebrate Paleontology* (ed. Thomason, J.) 219–234 (Cambridge Univ. Press, 1995).
83. Holland, S. M. *Paleobiology* **21**, 92–109 (1995).
84. Norell, M. A. in *Extinction and Phylogeny* (eds Novacek, M. J. & Wheeler, Q. D.) 89–118 (Columbia Univ. Press, New York, 1992).
85. Gradstein, F. M. et al. *Geophys. Res.* **99**, 24,051–24,074 (1994).
86. Cracraft, J. in *The Phylogeny and Classification of the Tetrapods Volume 1: Amphibians, Reptiles, Birds* (ed. Benton, M. J.) 339–361 (Clarendon, Oxford, 1988).
87. Norell, M. A. & Novacek, M. J. *Science* **255**, 1690–1693 (1992).

ACKNOWLEDGEMENTS. I thank A. Chinsamy, P. Dandonio, D. Frost, M. Norell, K. Padian, J. Wible and L. Witmer for comments and discussions. P. Conversano and M. Ellison prepared the illustrations. Research supported by the Frick Fund of the American Museum of Natural History, the Philip M. McKenna Foundation, and the National Science Foundation (DEB-9407999).

ARTICLES

A Jupiter-mass companion to a solar-type star

Michel Mayor & Didier Queloz

Geneva Observatory, 51 Chemin des Maillettes, CH-1290 Sauverny, Switzerland

The presence of a Jupiter-mass companion to the star 51 Pegasi is inferred from observations of periodic variations in the star's radial velocity. The companion lies only about eight million kilometres from the star, which would be well inside the orbit of Mercury in our Solar System. This object might be a gas-giant planet that has migrated to this location through orbital evolution, or from the radiative stripping of a brown dwarf.

FOR more than ten years, several groups have been examining the radial velocities of dozens of stars, in an attempt to identify orbital motions induced by the presence of heavy planetary companions^{1–5}. The precision of spectrographs optimized for Doppler studies and currently in use is limited to about 15 m s⁻¹. As the reflex motion of the Sun due to Jupiter is 13 m s⁻¹, all current searches are limited to the detection of objects with at least the mass of Jupiter (M_J). So far, all precise Doppler surveys have failed to detect any jovian planets or brown dwarfs.

Since April 1994 we have monitored the radial velocity of 142 G and K dwarf stars with a precision of 13 m s⁻¹. The stars in our survey are selected for their apparent constant radial velocity (at lower precision) from a larger sample of stars monitored for 15 years^{6,7}. After 18 months of measurements, a small number of stars show significant velocity variations. Although most candidates require additional measurements, we report here the discovery of a companion with a minimum mass of 0.5 M_J , orbiting at 0.05 AU around the solar-type star 51 Peg. Constraints originating from the observed rotational velocity of 51 Peg and from its low chromospheric emission give an upper limit of 2 M_J for

the mass of the companion. Alternative explanations to the observed radial velocity variation (pulsation or spot rotation) are unlikely.

The very small distance between the companion and 51 Peg is certainly not predicted by current models of giant planet formation⁸. As the temperature of the companion is above 1,300 K, this object seems to be dangerously close to the Jeans thermal evaporation limit. Moreover, non-thermal evaporation effects are known to be dominant⁹ over thermal ones. This jovian-mass companion may therefore be the result of the stripping of a very-low-mass brown dwarf.

The short-period orbital motion of 51 Peg also displays a long-period perturbation, which may be the signature of a second low-mass companion orbiting at larger distance.

Discovery of Jupiter-mass companion(s)

Our measurements are made with the new fibre-fed echelle spectrograph ELODIE of the Haute-Provence Observatory, France¹⁰. This instrument permits measurements of radial velocity with an accuracy of about 13 m s⁻¹ of stars up to 9 mag in an exposure time of <30 min. The radial velocity is computed

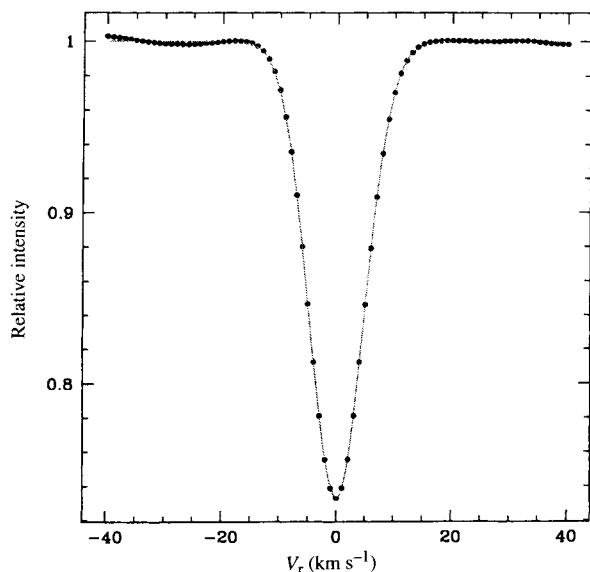


FIG. 1 Typical cross-correlation function used to measure the radial velocity. This function represents a mean of the spectral lines of the star. The location of the gaussian function fitted (solid line) is a precise measurement of the Doppler shift.

with a cross-correlation technique that concentrates the Doppler information of about 5,000 stellar absorption lines. The position of the cross-correlation function (Fig. 1) is used to compute the radial velocity. The width of the cross-correlation function is related to the star's rotational velocity. The very high radial-velocity accuracy achieved is a result of the scrambling effect of the fibres, as well as monitoring by a calibration lamp of instrumental variations during exposure.

The first observations of 51 Peg started in September 1994. In January 1995 a first 4.23-days orbit was computed and confirmed by intensive observations during eight consecutive nights in July 1995 and eight in September 1995. Nevertheless, a 24 m s^{-1} scatter of the orbital solution was measured. As this is incompatible with the accuracy of ELODIE measurements, we adjusted an orbit to four sets of measurements carried out at four different epochs with only the γ -velocity as a free parameter (see Fig. 2).

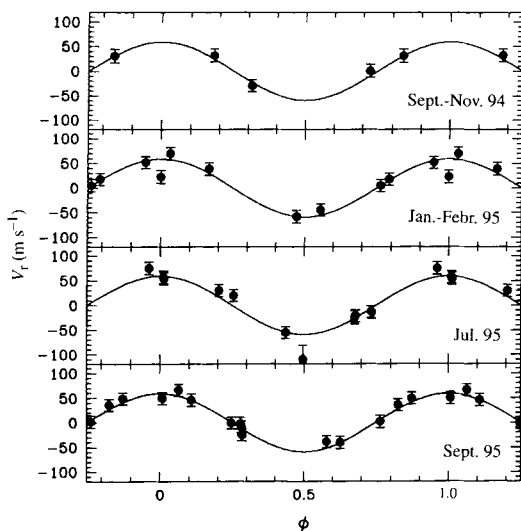


FIG. 2 Orbital motion of 51 Peg at four different epochs corrected from the γ -velocity. The solid line represents the orbital motion fitted on each time span with only the γ -velocity as a free parameter and with the other fixed parameters taken from Table 1.

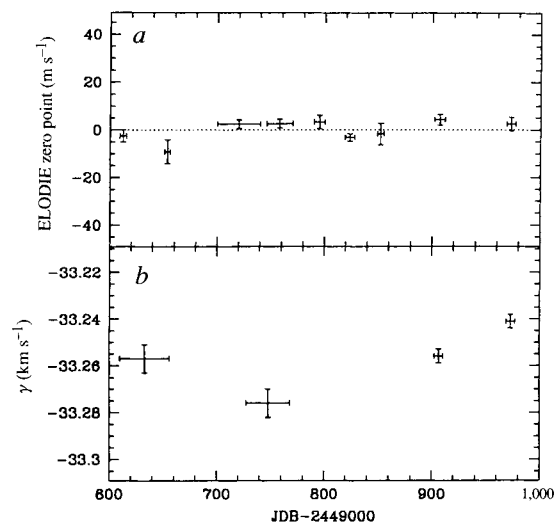


FIG. 3 a, ELODIE zero point computed from 87 stars of the sample having more than two measurements and showing no velocity variation. No instrumental zero point drift is detected. b, Variation of the γ -velocity of 51 Peg computed from the orbital fits displayed in Fig. 2. Considering the long-term stability of ELODIE this perturbation is probably due to a low-mass companion.

The γ -velocity in Fig. 3 shows a significant variation that cannot be the result of instrumental drift in the spectrograph. This slow perturbation of the short-period orbit is probably the signature of a second low-mass companion.

The long-period orbit cannot have a large amplitude. The 26 radial velocity measurements made during >12 years with the CORAVEL spectrometer do not reveal any significant variation at a 200 m s^{-1} level. Intensive monitoring of 51 Peg is in progress to confirm this long-period orbit.

In Fig. 4 a short-period circular orbit is fitted to the data after correction of the variation in γ -velocity. Leaving the eccentricity as a free parameter would have given $e = 0.09 \pm 0.06$ with almost the same standard deviation for the r.m.s. residual (13 m s^{-1}). Therefore we consider that a circular orbit cannot be ruled out. At present the eccentricity range is between 0 and about 0.15. Table 1 lists the orbital parameters of the circular-orbit solution.

An orbital period of 4.23 days is rather short, but short-period binaries are not exceptional among solar-type stars. (Five spectroscopic binaries have been found with a period <4 days in a volume-limited sample of 164 G-type dwarfs in the solar vicinity⁶.) Although this orbital period is not surprising in binary stars, it is puzzling when we consider the mass obtained for the companion:

$$M_2 \sin i = 0.47 \pm 0.02 M_J$$

where i is the (unknown) inclination angle of the orbit.

51 Peg (HR8729, HD217014 or Gliese 882) is a 5.5 mag star, quite similar to the Sun (see Table 2), located 13.7 pc (45 light yr) away. Photometric and spectroscopic analyses indicate a star slightly older than the Sun, with a similar temperature and slight overabundance of heavy elements. The estimated age¹¹ derived from its luminosity and effective temperature is typical of an old galactic-disk star. The slight overabundance of heavy elements in such an old disk star is noteworthy. But this is certainly not a remarkable peculiarity in view of the observed scatter of stellar metallicities at a given age.

Upper limit for the companion mass

A priori, we could imagine that we are confronted with a normal spectroscopic binary with an orbital plane almost perpendicular to the line of sight. Assuming a random distribution of binary orbital planes, the probability is less than 1% that the companion mass is larger than $4 M_J$, and $1/40,000$ that it is above the hydr-

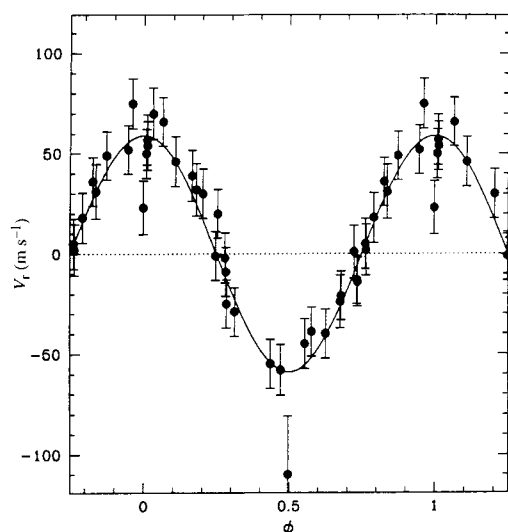


FIG. 4 Orbital motion of 51 Peg corrected from the long-term variation of the γ -velocity. The solid line represents the orbital motion computed from the parameters of Table 1.

ogen-burning limit of $0.08 M_{\odot}$. Although these probability estimates already imply a low-mass companion for 51 Peg, an even stronger case can be made from considerations of rotational velocity. If we assume that the rotational axis of 51 Peg is aligned with the orbital plane, we can derive $\sin i$ by combining the observed projected rotational velocity ($v \sin i$) with the equatorial velocity $V_{\text{equ}} = 2\pi R/P$ ($v \sin i = V_{\text{equ}} \cdot \sin i$).

Three independent precise $v \sin i$ determinations of 51 Peg have been made: by line-profile analysis¹², $v \sin i = 1.7 \pm 0.8 \text{ km s}^{-1}$; by using the cross-correlation function obtained with the CORAVEL spectrometer¹³, $v \sin i = 2.1 \pm 0.6 \text{ km s}^{-1}$; and by using the cross-correlation function obtained with ELODIE, $v \sin i = 2.8 \pm 0.5 \text{ km s}^{-1}$. The unweighted mean $v \sin i$ is $2.2 \pm 0.3 \text{ km s}^{-1}$. The standard error is probably not significant as the determination of very small $v \sin i$ is critically dependent on the supposed macroturbulence in the atmosphere. We accordingly prefer to admit a larger uncertainty: $v \sin i = 2.2 \pm 1 \text{ km s}^{-1}$.

51 Peg has been actively monitored for variability in its chromospheric activity¹⁴. Such activity, measured by the re-emission in the core of the Ca II lines, is directly related to stellar rotation via its dynamo-generated magnetic field. A very low level of chromospheric activity is measured for this object. Incidentally, this provides an independent estimate of an age of 10 Gyr (ref. 14), consistent with the other estimates. No rotational modulation has been detected so far from chromospheric emission, but a 30-day period is deduced from the mean chromospheric activity level S-index. A V_{equ} value of $2.2 \pm 0.8 \text{ km s}^{-1}$ is then com-

TABLE 2 Physical parameters of 51 Peg compared with those of the Sun

	51 Peg			
	Sun	Geneva photometry*	Spectroscopy†	Strömgren photometry and spectroscopy ¹¹
T_{eff} (K)	5,780	5,773	5,724	5,775
$\log g$	4.45	4.32	4.30	4.18
Fe/H	0		0.19	0.06‡
M/H	0	0.20		
M_V	4.79	4.60		
$R/R_{\odot}$	1	1.29		

M/H is the logarithmic ratio of the heavy element abundance compared to the Sun (in dex).

* M. Grenon (personal communication).

† J. Valenti (personal communication).

‡ But other elements such as Na I, Mg I, Al I are overabundant, in excess of 0.20.

puted if a 25% uncertainty in the period determination is assumed.

Using the mean $v \sin i$ and the rotational velocity computed from chromospheric activity, we finally deduce a lower limit of 0.4 for $\sin i$. This corresponds to an upper limit for the mass of the planet of $1.2 M_J$. Even if we consider a misalignment as large as 10° , the mass of the companion must still be less than $2 M_J$, well below the mass of brown dwarfs.

The 30-day rotation period of 51 Peg is clearly not synchronized with the 4.23-day orbital period of its low-mass companion, despite its very short period. (Spectroscopic binaries with similar periods are all synchronized.) The lack of synchronism on a timescale of 10^{10} yr is a consequence of the q^{-2} ($q = M_2/M_1$) dependence of the synchronization timescale¹⁵. In principle this can be used to derive an upper limit to the mass of the companion. It does at least rule out the possibility of the presence of a low-mass stellar companion.

Alternative interpretations?

With such a small amplitude of velocity variation and such a short period, pulsation or spot rotation might explain the observations equally well^{16,17}. We review these alternative interpretations below and show that they can probably be excluded.

Spot rotation can be dismissed on the basis of the lack of chromospheric activity and the large period derived from the S chromospheric index, which is clearly incompatible with the observed radial-velocity short period. A solar-type star rotating with a period of 4.2 days would have a much stronger chromospheric activity than the currently observed value¹⁴. Moreover, a period of rotation of 4.2 days for a solar-type star is typical of a very young object (younger than the Pleiades) and certainly not of an old disk star.

Pulsation could easily yield low-amplitude velocity variations similar to the one observed, but would be accompanied by luminosity and colour variations as well as phase-related absorption line asymmetries. The homogeneous photometric survey made by the Hipparcos satellite provides a comprehensive view of the intrinsic variability of stars of different temperatures and luminosities. The spectral type of 51 Peg corresponds to a region of the Hertzsprung–Russell diagram where the stars are the most stable¹⁸.

Among solar-type stars no mechanisms have been identified for the excitation of pulsation modes with periods as long as 4 days. Only modes with very low amplitude ($\ll 1 \text{ m s}^{-1}$) and periods from minutes to 1 h are detected for the Sun.

Radial velocity variations of a few days and $< 100 \text{ m s}^{-1}$ amplitude have been reported for a few giant stars¹⁹. Stars with a similar spectral type and luminosity class are known to be photometric variables¹⁸. Their observed periods are in

TABLE 1 Orbital parameters of 51 Peg

P	$4.2293 \pm 0.0011 \text{ d}$
T	$2,449,797.773 \pm 0.036$
e	0 (fixed)
K_1	$0.059 \pm 0.003 \text{ km s}^{-1}$
$a_1 \sin i$	$(34 \pm 2) 10^5 \text{ m}$
$f_1(m)$	$(0.91 \pm 0.15) 10^{-10} M_{\odot}$
N	35 measurements
(O–C)	13 m s^{-1}

P , period; T , epoch of the maximum velocity; e , eccentricity; K_1 , half-amplitude of the velocity variation; $a_1 \sin i$, where a_1 is the orbital radius; $f_1(m)$, mass function; N , number of observations; (O–C), r.m.s. residual.

agreement with predicted pulsation periods for giant stars with radii $>20 R_{\odot}$. 51 Peg, with its small radius, can definitely not be compared to these stars. These giant stars also pulsate simultaneously in many short-period modes, a feature certainly not present in the one-year span of 51 Peg observations. It is worth noticing that 51 Peg is too cold to be in the δ Scuti instability strip.

G. Burki *et al.* (personal communication) made 116 photometric measurements of 51 Peg and two comparison stars in the summer of 1995 at ESO (La Silla) during 17 almost-consecutive nights. The observed magnitude dispersions for the three stars are virtually identical, respectively $V=0.0038$ for 51 Peg, and $V=0.0036$ and 0.0039 for the comparison stars. The fit of a sine curve with a period of 4.2293 days to the photometric data limits the possible amplitude to 0.0019 for V magnitude and 0.0012 for the $[B_2 - V_1]$ Geneva colour index. Despite the high precision of these photometric measurements we cannot completely rule out, with these photometric data alone, the possibility of a very low-amplitude pulsation. In the coming months, stronger constraints can be expected from the numerous Hipparcos photometric data of this star.

Pulsations are known to affect the symmetry of stellar absorption lines. To search for such features we use the cross-correlation technique, as this technique is a powerful tool for measuring mean spectral line characteristics²⁰. The difference in radial velocity of the lower and upper parts of the cross-correlation function is an indicator of the line asymmetry. The amplitude of a 4.2-day sine curve adjusted to this index is less than 2 m s^{-1} . The bisector of the cross-correlation function does not show any significant phase variation.

From all the above arguments, we believe that the only convincing interpretation of the observed velocity variations is that they are due to the orbital motion of a very-low-mass companion.

Jupiter or stripped brown dwarf?

At the moment we certainly do not have an understanding of the formation mechanism of this very-low-mass companion. But we can make some preliminary comments about the importance of evaporation as well as the dynamic evolution of the orbit.

If we compare 51 Peg b with other planets or G-dwarf stellar companions (Fig. 5) it is clear that the mass and the low orbital eccentricity of this object are in the range of heavy planets, but this certainly does not imply that the formation mechanism of this planet was the same as for Jupiter.

Present models for the formation of Jupiter-like planets do not allow the formation of objects with separations as small as 0.05 AU. If ice grains are involved in the formation of giant planets, the minimum semi-major axis for the orbits is about 5 AU (ref. 8), with a minimum period of the order of 10 yr. A Jupiter-type planet probably suffers some orbital decay during its formation by dynamic friction. But it is not clear that this could produce an orbital shrinking from 5 AU to 0.05 AU.

All of the planets in the Solar System heavier than $10^{-6} M_{\odot}$ have almost circular orbits as a result of their origin from a protoplanetary gaseous disk. Because of its close separation, however, the low eccentricity of 51 Peg b is not a proof of similar origin. Tidal dissipation acting on the convective envelope is known¹⁵ to circularize the orbit and produce a secular shrinking of the semi-major axis of binary systems. The characteristic time is essentially proportional to $q^{-1} P^{16/3}$. For stars of the old open cluster M67, orbital circularization is observed for periods lower than 12.5 days (ref. 21). We derive for 51 Peg a circularization time of a few billion years, shorter than the age of the system. The low orbital eccentricity of 51 Peg b could result from the dynamic evolution of the system and not necessarily from its formation conditions.

A Jupiter-sized planet as close as 0.05 AU to 51 Peg should have a rather high temperature of about 1,300 K. To avoid a significant evaporation of a gaseous atmosphere, the escape

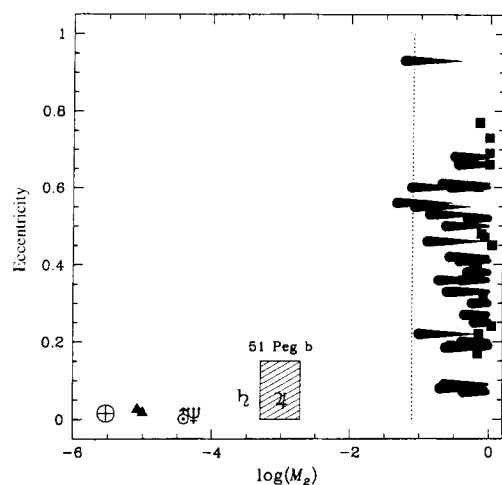


FIG. 5 Orbital eccentricities of planets as well as companion of G-dwarf binaries⁸ in the solar vicinity as a function of their mass M_2 . The planets of the Solar System are indicated with their usual symbols. The planets orbiting around the pulsar^{24,25} PSR B 1257+12 are indicated by filled triangles. The uncertainties on the mass of SB1 (single-spectrum spectroscopic binaries), owing to their unknown orbital inclination, are indicated by an elongated line that thins to a $\sin i$ probability of 99%. SB2s are indicated by filled squares. (Only the stellar orbits not tidally circularized with periods larger than 11 days are indicated.) Note the discontinuity in the orbital eccentricities when planets are binary stars are compared, and the gap in masses between the giant planets and the lighter secondaries of solar-type stars. The dotted line at $0.08 M_{\odot}$ indicates the location of the minimum mass for hydrogen burning. The position of 51 Peg b with its uncertainties is indicated by the hatched rectangle.

velocity V_e has to be larger than the thermal velocity V_{th} : $V_e > \alpha V_{th}$. This imposes a minimum mass for a gaseous planet at a given separation:

$$\frac{M_p}{M_J} > \alpha^2 \left(\frac{kT_*}{m} \right) \left(\frac{GM_J}{R_p} \right)^{-1} (1-\gamma)^{1/4} \left(\frac{R_*}{2a} \right)^{1/2}$$

where γ denotes the albedo of the planet, R_p and M_p are its radius and mass, m is the mass of atoms in the planet atmosphere, and R_* and T_* are the radius and effective temperature of the star.

Our lack of knowledge of the detailed structure of the atmosphere of the planet prevents us from making an accurate estimate of α . A first-order estimate of $\alpha \approx 5-6$ is nevertheless made by analogy with planets of the Solar System²². We find that with a planetary radius probably increased by a factor of 2-3 owing to the high surface temperature (A. Burrows, personal communication), gaseous planets more massive than $0.6-1.0 M_J$ are at the borderline for suffering thermal evaporation. Moreover, for the Solar-System planets, non-thermal evaporative processes are known to be more efficient than thermal ones⁹. The atmosphere of 51 Peg b has thus probably been affected by evaporation.

Recent work²³ on the fragmentation of molecular clouds shows that binary stars can be formed essentially as close to each other as desired, especially if the effects of orbital decay are considered. We can thus speculate that 51 Peg b results from a strong evaporation of a very close and low-mass brown dwarf. In such a case 51 Peg b should mostly consist of heavy elements. This model is also not free of difficulties, as we expect that a brown dwarf suffers less evaporation owing to its larger escape velocity.

We are eager to confirm the presence of the long-period companion and to find its orbital elements. If its mass is in the range of a few times that of Jupiter and its orbit is also quasi-circular, 51 Peg could be the first example of an extrasolar planetary system associated with a solar-type star.

The search for extrasolar planets can be amazingly rich in surprises. From a complete planetary system detected around a pulsar^{24,25}, to the rather unexpected orbital parameters of 51 Peg b, searches begin to reveal the extraordinary diversity of possible planetary formation sites.

Note added in revision: After the announcement of this discovery at a meeting held in Florence, independent confirmations of

the 4.2-day period radial-velocity variation were obtained in mid-October by a team at Lick Observatory, as well as by a joint team from the High Altitude Observatory and the Harvard-Smithsonian Center for Astrophysics. We are deeply grateful to G. Marcy, P. Butler, R. Noyes, T. Kennelly and T. Brown for having immediately communicated their results to us. □

Received 29 August; accepted 31 October 1995.

- Walker, G. A. H., Walker, A. R. & Irwin, A. W. *Icarus* **116**, 359–375 (1995).
- Cochran, W. D. & Hatzes, A. P. *Astrophys. Space Sci.* **212**, 281–291 (1994).
- Marcy, G. W. & Butler, R. P. *Publ. astr. Soc. Pacif.* **104**, 270–277 (1992).
- McMillan, R. S., Moore, T. L., Perry, M. L. & Smith, P. H. *Astrophys. Space Sci.* **212**, 271–280 (1994).
- Marcy, G. W. & Butler, R. P. in *The Bottom of the Main Sequence and Beyond* (ESO Astrophys. Symp.) (ed. Tinney, C. G.) 98–108 (Springer, Berlin, 1995).
- Duquennoy, A. & Mayor, M. *Astr. Astrophys.* **248**, 485–524 (1991).
- Mayor, M., Duquennoy, A., Halbwachs, J. L. & Mermilliod, J. C. in *Complementary Approaches to Double and Multiple Star Research* (eds McAlister, A. A. & Hartkopf, W. I.) (ASP Conf. Ser. 32, 73–81 (Astr. Soc. Pacific, California, 1992).
- Boss, A. P. *Science* **267**, 360–362 (1995).
- Hunten, D. H., Donahue, T. M., Walker, J. C. G. & Kasting, J. F. in *Origin and Evolution of Planetary and Satellite Atmospheres* (eds Atreya, S. K., Pollack, J. B. & Matthews, M. S.) 386–422 (Univ. of Arizona Press, Tucson, 1989).
- Baranne, A. *Astrophys. J. Suppl.* (submitted).
- Edvardsson, B. et al. *Astr. Astrophys.* **275**, 101–152 (1993).
- Soderblom, D. R. *Astrophys. J. Suppl. Ser.* **53**, 1–15 (1983).
- Baranne, A., Mayor, M. & Poncet, J. L. *Vistas Astr.* **23**, 279–316 (1979).
- Noyes, R. W., Hartmann, L. W., Baliunas, S. L., Duncan, D. K. & Vaughan, A. H. *Astrophys. J.* **279**, 763–777 (1984).

- Zhan, J. P. *Astr. Astrophys.* **220**, 112–116 (1989).
- Walker, G. A. H. et al. *Astrophys. J.* **396**, L91–L94 (1992).
- Larson, A. M. et al. *Publ. astr. Soc. Pacif.* **105**, 825–831 (1993).
- Eyer, L., Grenon, M., Falin, J. L., Froeschlé, M. & Mignard, F. *Sol. Phys.* **152**, 91–96 (1994).
- Hatzes, A. P. & Cochran, W. D. *Proc. 9th Cambridge Workshop* (ed. Pallavicini, R.) (Astronomical Soc. of the Pacific) (in the press).
- Queloz, D. in *New Developments in Array Technology and Applications* (eds Davis Philip, A. G. et al.) 221–229 (Int. Astr. Union, 1995).
- Latham, D. W., Mathieu, R. D., Milone, A. A. E. & Davis, R. J. in *Binaries as Tracers of Stellar Formation* (eds Duquennoy, A. & Mayor, M.) 132–138 (Cambridge Univ. Press, 1992).
- Lewis, J. S. & Prinn, R. G. *Planets and their Atmospheres—Origin and Evolution* (Academic, Orlando, 1984).
- Bonnell, I. A. & Bate, M. R. *Mon. Not. R. astr. Soc.* **271**, 999–1004 (1994).
- Wolszczan, A. & Frail, D. A. *Nature* **355**, 145–147 (1992).
- Wolszczan, A. *Science* **264**, 538–542 (1994).

ACKNOWLEDGEMENTS. We thank G. Burki for analysis of photometric data, W. Benz for stimulating discussions, A. Burrows for communicating preliminary estimates of the radius of Jupiter at different distances from the Sun, and F. Pont for his careful reading of the manuscript. We also thank all our colleagues of Marseille and Haute-Provence Observatories involved in the building and operation of the ELODIE spectrograph, namely G. Adrianszyk, A. Baranne, R. Cautain, G. Knispel, D. Kohler, D. Lacroix, J.-P. Meunier, G. Rimbaud and A. Vin.

LETTERS TO NATURE

Unusually strong intramolecular magnetic coupling in a chromium hydride cluster

Robert A. Heintz*, Thomas F. Koetzle†, Robert L. Ostrander*, Arnold L. Rheingold*, Klaus H. Theopold*† & Ping Wu†

* Department of Chemistry and Biochemistry, University of Delaware, Newark, Delaware 19716, USA

† Department of Chemistry, Brookhaven National Laboratory, Upton, New York 11973, USA

MOLECULES with large spins, and associated large magnetic moments, are potential building blocks for magnetic materials¹. To make such molecules, spin-carrying metal ions can be assembled into polynuclear complexes, but the coupling between the spins is usually antiferromagnetic, leading to antiparallel alignment². Magnetic coupling leading to parallel alignment occurs only rarely, and even then is usually too weak for the alignment to persist at ambient temperatures^{3–14}. Here we report the synthesis of a tetranuclear chromium hydride cluster with a ground state of non-zero spin (spin quantum number $S = 7/2$), in which the intramolecular magnetic coupling is so strong that the magnetic alignment is not disturbed appreciably even at room temperature. This ground state cannot be explained either by simple parallel or antiparallel alignment of spins, but can be understood in terms of antiparallel alignment of three Cr(III) moments with one Cr(II) moment. These findings indicate that hydride ligands can mediate extremely strong magnetic exchange interactions between metal ions, and that metal hydrides may therefore be promising components for the construction of molecular magnetic materials.

Exposure of chromium(II) hydride [$\text{Cp}^*\text{Cr}_4(\mu_3\text{-H})_4$] (where Cp^* is η^5 -tetramethyl-ethyl-cyclopentadienyl)¹⁵ to hydrogen gas

at high pressure (60 atm, C_5H_{12} solvent, 25 °C, 5 days) yielded a new compound, tentatively identified as the polyhydride [$\text{Cp}^*\text{Cr}_4\text{H}_n$], in 65% isolated yield. The substance was highly paramagnetic; thus its ^1H NMR spectrum (in C_6D_6) exhibited only two very broad and isotropically shifted resonances (chemical shift 9.0 p.p.m., full width at half maximum (FWHM) 570 Hz and –0.9 p.p.m., FWHM 280 Hz) attributable to the substituents of the cyclopentadienyl ligands. No hydride resonances were found. A single-crystal X-ray diffraction study yielded the positions of the heavy atoms ($T = 298$ K, orthorhombic, space group $Pnma$, unit-cell constants $a = 15.447$ (3) Å, $b = 17.242$ (5) Å, $c = 16.221$ (5) Å, $V = 4320.5$ (21) Å³, $Z = 4$, $d_{\text{calc}} = 1.250$ g cm^{–3}, $\mu = 1.012$ mm^{–1}, (where Z is the number of molecules per unit cell, d_{calc} is the calculated density and μ is the linear absorption coefficient) 4,989 reflections collected, 1,657 observed ($F > 5.0\sigma(F)$), $R = 5.40\%$, $R_w = 6.16\%$). Like its precursor, the new cluster consists of an approximately tetrahedral array of four chromium atoms, each of which is bonded to an alkyl-substituted cyclopentadienyl ligand in the familiar η^5 -fashion. But in contrast to the tetrahydride, this compound lies on a crystallographic mirror plane, and its Cr–Cr distances (Cr–Cr_{avg} = 2.765 Å, (where Cr_{avg} is the average Cr–Cr distance) range 2.759 (3) to 2.767 (2) Å) are significantly longer (Cr–Cr_{avg} = 2.651 Å for [$\text{Cp}^*\text{Cr}_4(\mu_3\text{-H})_4$]). The latter observation is consistent with an increase in the formal oxidation state of chromium, as could be brought about by oxidative addition of H_2 to the starting material. The X-ray structure determination revealed several peaks of residual electron density, which might indicate hydride positions, but an unambiguous determination of the number and positions of chromium-bonded hydrogen atoms proved impossible.

Under these circumstances, a neutron-diffraction study was required to unambiguously characterize the new polyhydride. Diffraction data were acquired at the Brookhaven High Flux Beam Reactor on a large single crystal (3.3 × 4.3 × 5.4 mm). The initial atomic parameters were taken from the X-ray analysis, and the hydrogen atoms were located from successive Fourier-difference maps. The result of the structure determination is shown in Fig. 1 and selected interatomic distances and angles are listed in Table 1 (temperature $T = 20$ K, orthorhombic, space

† To whom correspondence should be addressed.

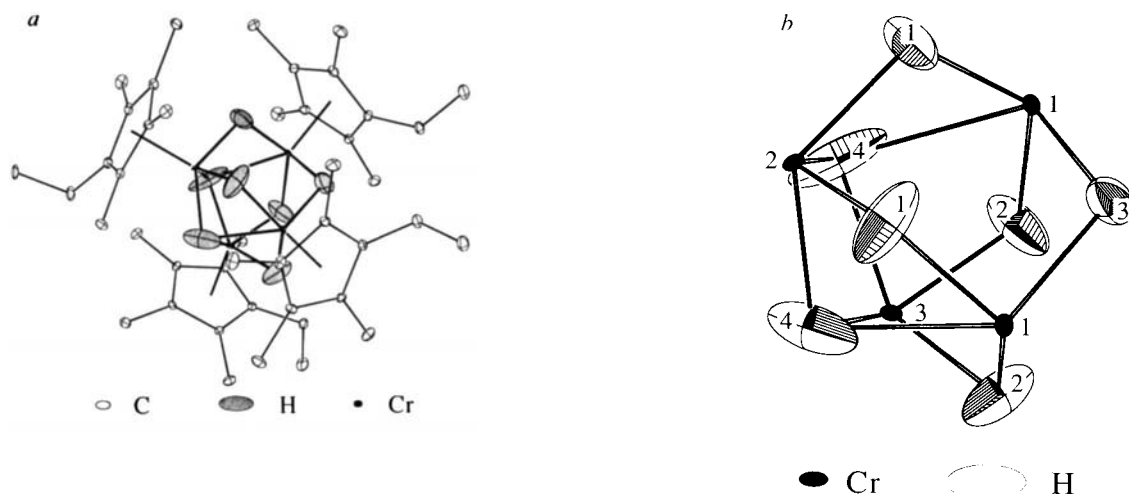


FIG. 1 a, The structural framework, and b, the chromium hydride core of $[\text{Cp}^*_4\text{Cr}_4(\mu_2\text{-H})_5(\mu_3\text{-H})_2]$ with the atom numbering scheme. The

group $Pnma$, unit-cell constants $a=15.347(18)$ Å, $b=17.051(8)$ Å, $c=15.809(7)$ Å, $V=4137(6)$ Å³, $Z=4$, $d_{\text{calc}} 1.304$ g cm⁻³, $\mu=0.302$ mm⁻¹, 5,552 unique data, 3,050 observed ($F_o^2 > 3\sigma(F_o^2)$, $R=11.7\%$, $R_w=9.8\%$). The heavy-atom positions substantially agree with the X-ray diffraction results, although there are slight changes in bond distances and angles due to the large temperature difference. However, the neutron-diffraction experiment also revealed seven hydride ligands, thus enabling the unambiguous assignment of the formula $[\text{Cp}^*_4\text{Cr}_4(\mu_2\text{-H})_5(\mu_3\text{-H})_2]$ to the new cluster. The core of the molecule (see Fig. 1b) is best described as a tetrahedron of metal atoms, with all but one of its six edges bridged by one hydride ligand each, while the sixth edge (that is, Cr(2)–Cr(3)) supports two bridging hydrides. The latter are tilted somewhat toward chromium atoms Cr(1) and Cr(1a), respectively, thereby approaching a face-bridging ($\mu_3\text{-H}$) position. The Cr–H distances average 1.733 Å (range 1.696 (8) to 1.757 (7) Å), very similar to the corresponding values in $[\text{Cr}_2(\text{CO})_{10}(\mu_2\text{-H})]^-$ (ref. 16), while the long Cr(1)–H(4) distance measures 2.218 (10) Å. $[\text{Cp}^*_4\text{Cr}_4(\mu_2\text{-H})_5(\mu_3\text{-H})_2]$ is a mixed-valent complex containing three chromium atoms in the formal oxidation state III (d^3 , $S=3/2$) and one chromium(II) (d^4 , $S=1$). However, neither the molecular structure nor the individual bond distances allow the identification of a unique divalent chromium.

TABLE 1 Selected interatomic distances and angles for $[\text{Cp}^*_4\text{Cr}_4(\mu_2\text{-H})_5(\mu_3\text{-H})_2]$

Distances (Å)			
Cr(1)–Cr(1a)	2.803 (7)	Cr(1)–Cr(2)	2.769 (6)
Cr(1)–Cr(3)	2.787 (5)	Cr(2)–Cr(3)	2.699 (7)
Cr(1)–H(1)	1.737 (7)	Cr(1)–H(2)	1.753 (6)
Cr(1)–H(3)	1.729 (5)	Cr(1)–H(4)	2.218 (10)
Cr(2)–H(1)	1.696 (7)	Cr(2)–H(4)	1.749 (7)
Cr(3)–H(2)	1.712 (7)	Cr(3)–H(4)	1.757 (7)
Angles (degrees)			
H(3)–Cr(1)–H(1)	97.2 (4)	H(3)–Cr(1)–H(2)	94.8 (4)
H(3)–Cr(1)–H(4)	118.0 (3)	H(1)–Cr(1)–H(2)	127.7 (3)
H(1)–Cr(1)–H(4)	64.3 (3)	H(2)–Cr(1)–H(4)	65.1 (3)
H(1)–Cr(2)–H(1a)	82.8 (6)	H(1)–Cr(2)–H(4)	77.0 (4)
H(1)–Cr(2)–H(4a)	77.8 (4)	H(2)–Cr(3)–H(2a)	80.4 (6)
H(2)–Cr(3)–H(4)	77.5 (4)	H(2)–Cr(3)–H(4a)	126.3 (4)
H(4)–Cr(3)–H(4a)	77.4 (5)	Cr(1)–H(1)–Cr(2)	107.6 (4)
Cr(1)–H(2)–Cr(3)	107.1 (4)	Cr(1)–H(3)–Cr(1a)	108.4 (4)
Cr(2)–H(4)–Cr(3)	100.6 (3)	Cr(1)–H(4)–Cr(2)	87.7 (4)
Cr(1)–H(4)–Cr(3)	88.3 (4)		

hydrogen atoms of the alkyl groups were omitted for clarity. The ellipsoids are shown at the 50% probability level.

One of the few known chromium hydrides^{16, 21}, the molecule reported here also belongs to the small number of paramagnetic hydride complexes containing more than one metal atom^{22–27}. Thus it provides an opportunity to investigate the efficacy of hydrogen atoms in mediating magnetic interactions between metal atoms with unpaired spins. The short metal–metal distances enforced by bridging hydride ligands, and the intervention of a single, radially symmetric and large ligand orbital (the hydride ion has a $1s^2$ electronic configuration and an ionic radius of 2.08 Å) between the magnetic ions, would be expected to enhance any magnetic exchange interaction by maximizing overlap of the relevant orbitals⁴. By analogy to $[\text{Cp}^*_4\text{Cr}_4(\mu_3\text{-H})_4]$, which features antiferromagnetic coupling between its four Cr^{II} ions ($S=1$)¹⁵, the heptahydride might have been expected to exhibit a relatively low effective magnetic moment, and a decrease of the latter with temperature. However, its magnetic behaviour is highly unusual. The molar magnetic susceptibility of the cluster (χ_m) was measured with a Faraday balance ($H=9.6 \times 10^3$ A m⁻¹) in the temperature interval 60–300 K. The temperature dependence of χ_m , and of the effective magnetic moment (μ_{eff}) calculated therefrom, is depicted in Fig. 2. The hydride cluster exhibited nearly ideal Curie behavior, and a fit with a modified Curie–Weiss expression ($\chi_m = C/(T - \theta) + \text{TIP}$; where T is the absolute temperature, C is the Curie constant and TIP is temperature independent magnetism) yielded a small Weiss constant ($\theta = -3.0(2)$ K). Accordingly, the effective magnetic moment was independent of temperature and measured $\mu_{\text{eff}} = 8.1 \mu_B$ where μ_B is the Bohr magneton.

The magnetic behaviour of mixed-valent $[\text{Cp}^*_4\text{Cr}_4(\mu_2\text{-H})_5(\mu_3\text{-H})_2]$ was obviously not consistent with antiferromagnetic coupling, which would result in a molecular ground state with $S=1/2$. Two alternatives had to be considered. However unlikely, considering the short metal–metal distances and multiple bridging ligands, the metal ions might be magnetically independent of each other. Summation of the individual susceptibility contributions of three Cr(III) ions and one Cr(II) ion, followed by calculation of the expected effective magnetic moment per Cr₄ cluster, gave $\mu_{\text{eff}} = 7.28 \mu_B$, that is, of the same order of magnitude as the measured value. On the other hand, although a straightforward ferromagnetic coupling of all four spins (ground state $S=11/2$) would result in too high an expected moment ($\mu_{\text{eff}} = 11.96 \mu_B$), more complicated interactions are also possible. For example, antiparallel alignment of all $S=3/2$ spins with the unique $S=1$ spin would result in an overall spin state of $S=7/2$. Such a situation essentially represents a molecular version of ferrimagnetism. The spin-only

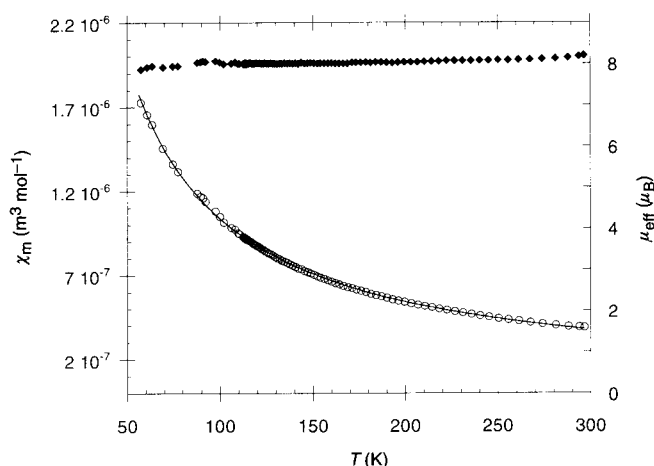


FIG. 2 Molar magnetic susceptibility (χ_m , $\circ$, $1 \text{ emu mol}^{-1} = (1/4\pi) \times 10^6 \text{ m}^3 \text{ mol}^{-1}$) and effective magnetic moment (μ_{eff} , $\blacklozenge$) of $[\text{Cp}_4\text{Cr}_4(\mu_2\text{-H})_5(\mu_3\text{-H})_2]$ in the temperature interval 60–300 K. The susceptibility data were checked for ferromagnetic impurities (none apparent) by the Honda-Owens method³³, and corrected for diamagnetism ($-7.36 \times 10^{-9} \text{ m}^3 \text{ mol}^{-1}$). Line represents a fit of the data with a Curie-Weiss expression ($\chi_m = C/(T - \theta) + \text{TIP}$; $C = 1.04 \times 10^{-4} \text{ m}^3 \text{ K mol}^{-1}$, $\theta = -3.0(2) \text{ K}$, $\text{TIP} = 3.1(1) \times 10^{-8} \text{ m}^3 \text{ mol}^{-1}$).

moment expected of a molecule with $S = 7/2$ is $\mu_{\text{eff}} = 7.94 \mu_B$, very close to the experimentally determined value.

This ambiguity in interpretation was resolved by magnetization measurements. Figure 3 shows the magnetic-field dependence of the magnetization M of the cluster measured with a SQUID magnetometer at low temperatures. Fitting of the data with Brillouin functions (assuming a Landé constant (g) of 2.0) gave three independent values of the spin quantum number S of the ground state, averaging to $S = 3.4(2)$. This experiment thus supports the notion that $[\text{Cp}_4\text{Cr}_4(\mu_2\text{-H})_5(\mu_3\text{-H})_2]$ is a high-spin molecule with seven unpaired electrons ($S = 7/2$), resulting from intramolecular ferrimagnetic alignment as described above.

Parallel alignment of spins in polynuclear transition metal complexes is the exception. Even when observed, the weakness of the coupling usually restricts the exclusive occupation of high-spin states to very low temperatures³. Increasing temperature typically results in population of energetically higher spin states with lower spin multiplicities, thus leading to decreasing effective magnetic moments. Under these circumstances, analysis of the temperature dependence of χ_m or μ_{eff} according to the Heisenberg-Dirac-Van Vleck (HDDVV) model allows a determi-

nation of the energetic separation of the spin states (or the related coupling constant J)²⁸. Returning to Fig. 3, it may be seen that the effective moment of $[\text{Cp}_4\text{Cr}_4(\mu_2\text{-H})_5(\mu_3\text{-H})_2]$ shows no sign of a decrease with temperature up to 300 K. Apparently, the thermal energy available at room temperature ($\sim 200 \text{ cm}^{-1}$) is not sufficient to significantly populate any other spin states of the molecule. Although the absence of a temperature effect precludes a numerical determination of a coupling constant J , it implies that the value must be in excess of several hundred cm^{-1} . This represents, to our knowledge, the strongest parallel magnetic alignment yet observed in any molecular cluster.

Strong magnetic interactions between individual atomic spins are crucial for bulk magnetism at ambient temperatures. Selected metals (Fe, Co, Ni) and metal oxides (Fe_3O_4 , CrO_2) have long been recognized to be magnetic at room temperature or above. The recent search for molecular magnets has produced some materials with relatively high Curie temperatures (T_C), namely metal cyanides^{29–31} and the enigmatic³² $[\text{V}(\text{TCNE})_2] \cdot \frac{1}{2} \text{CH}_2\text{Cl}_2$ (where TCNE is tetracyanoethylene). Our results suggest that metal hydrides may also have potential as building blocks for new magnetic materials. □

Received 22 May; accepted 10 October 1995.

1. Miller, J. S. & Epstein, A. J. *Angew. Chem. int. Edn engl.* **33**, 385–415 (1994).
2. Cairns, C. J. & Busch, D. H. *Coord. Chem. Rev.* **69**, 1–55 (1986).
3. Bertrand, J. A. et al. *Inorg. Chem.* **10**, 240–246 (1971).
4. Kahn, O. *Angew. Chem. int. Edn engl.* **24**, 834–850 (1985).
5. Bino, A., Johnston, D. C., Goshorn, D. P., Halbert, T. R. & Stiefel, E. I. *Science* **241**, 1479–1481 (1988).
6. Caneschi, A. et al. *J. Am. chem. Soc.* **110**, 2795–2799 (1988).
7. Hendrickson, D. N. et al. *J. Am. chem. Soc.* **114**, 2455–2471 (1992).
8. Goldberg, D. P., Caneschi, A. & Lippard, S. J. *J. Am. chem. Soc.* **115**, 9299–9300 (1993).
9. Sessoli, R. et al. *J. Am. chem. Soc.* **115**, 1804–1816 (1993).
10. Blake, A. J., Grant, C. M., Parsons, S., Rawson, J. M. & Winpenny, R. E. P. *J. chem. Soc., chem. Commun.* 2363–2364 (1994).
11. Wang, Z. R., Wang, X.-L., Fernandez-Baca, J. A., Johnston, D. C. & Vukobratovic, D. *Science* **264**, 402–404 (1994).
12. Epley, H. J. et al. *J. Am. chem. Soc.* **117**, 301–317 (1995).
13. Goldberg, D. P., Caneschi, A., Delfs, C. D., Sessoli, R. & Lippard, S. J. *J. Am. chem. Soc.* **117**, 5789–5800 (1995).
14. Wemple, M. W., Adams, D. M., Folting, K., Hendrickson, D. N. & Christou, G. *J. Am. chem. Soc.* **117**, 7275–7276 (1995).
15. Heintz, R. A., Ostrander, R. L., Rheingold, A. L. & Theopold, K. H. *J. Am. chem. Soc.* **116**, 11387–11396 (1994).
16. Petersen, J. L., Brown, R. K. & Williams, J. M. *Inorg. Chem.* **20**, 158–165 (1981).
17. Fischer, E. O. & Hafner, W. Z. *Naturf.* **108**, 140–143 (1955).
18. Daresbourg, M. Y. et al. *J. Am. chem. Soc.* **104**, 6961–6969 (1984).
19. Girolami, G. S., Salt, J. E., Wilkinson, G., Thornton-Pett, M. & Hursthouse, M. B. *J. Am. chem. Soc.* **105**, 5954–5956 (1983).
20. Van-Cattledge, F. A., Ittel, S. D. & Jesson, J. P. *Organometallics* **4**, 18–27 (1985).
21. Thaler, E., Folting, K., Huffman, J. C. & Caulton, K. G. *Inorg. Chem.* **26**, 374–377 (1987).
22. Müller, J., Dörner, H., Huttner, G. & Lorenz, H. *Angew. Chem. int. Edn engl.* **12**, 1005–1006 (1973).
23. Huttner, G. & Lorenz, H. *Chem. Ber.* **107**, 996–1008 (1974).
24. Koetzle, T. F. et al. *Adv. Chem. Ser.* **167**, 61–72 (1978).
25. Samuel, E. & Harrod, J. F. *J. Am. chem. Soc.* **106**, 1859–1860 (1984).
26. Bianchini, C. et al. *Inorg. Chem.* **28**, 2552–2560 (1989).
27. Kersten, J. L. et al. *Angew. Chem. int. Edn engl.* **31**, 1341–1343 (1992).
28. Martin, R. L. in *New Pathways in Inorganic Chemistry* (eds Ebsworth, E. A. V., Maddock, A. G. & Sharpe, A. G.) 175–231 (Cambridge Univ. Press, 1968).

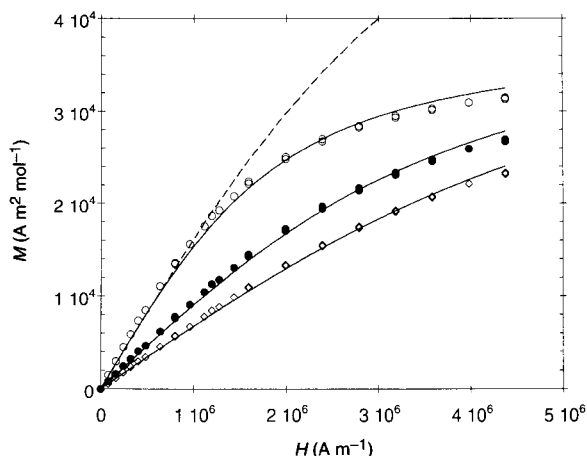


FIG. 3 Magnetic field (H , $1 \text{ Oe} = 4\pi \times 10^{-3} \text{ A m}^{-1}$) dependence of the magnetization (M , $1 \text{ emu mol}^{-1} = 1 \text{ A m}^2 \text{ mol}^{-1}$) of $[\text{Cp}_4\text{Cr}_4(\mu_2\text{-H})_5(\mu_3\text{-H})_2]$ at various low temperatures (5 K, $\circ$; 10 K, $\bullet$; 15 K, $\diamond$). Solid lines represent fits of the data with Brillouin functions. Fit parameter is the spin quantum number S ; at 5 K $S = 3.20$, at 10 K $S = 3.38$, at 15 K $S = 3.57$. Dashed line shows magnetization expected of a system of four non-interacting chromium ions (three $S = 3/2$ and one $S = 1$) at 5 K.

29. Gadet, V., Mallah, T., Castro, I. & Verdaguer, M. *J. Am. chem. Soc.* **114**, 9213–9214 (1992).
 30. Mallah, T., Thiebaut, S., Verdaguer, M. & Veillet, P. *Science* **262**, 1554–1557 (1993).
 31. Entley, W. R. & Girolami, G. S. *Science* **268**, 397–400 (1995).
 32. Manríquez, J. M., Yee, G. T., McLean, R. S., Epstein, A. J. & Miller, J. S. *Science* **252**, 1415–1417 (1991).
 33. Bates, L. F. *Modern Magnetism* 4th edn 133–136 (Cambridge Univ. Press, London, 1961).

ACKNOWLEDGEMENTS. We thank K. Unruh and L. Withanawasam for the magnetization measurements, and C. Koehler for technical assistance with the neutron-diffraction experiment. This work was supported by grants from the US NSF and DOE.

Noise-induced enhancement of signal transduction across voltage-dependent ion channels

Sergey M. Bezrukov*† & Igor Vodyanoy*†§

* Division of Computer Research and Technology,
National Institutes of Health, Bethesda,
Maryland 20892-0580, USA

† St Petersburg Nuclear Physics Institute, Gatchina, Russia 188350

‡ Office of Naval Research, Europe, 223 Old Marylebone Road,
London NW1 5TH, UK

§ University College London, Gower Street, London, UK

THE presence of noise in a signal transduction system usually interferes with its ability to transfer information reliably. But many nonlinear systems can use noise to enhance performance¹, and this phenomenon, called stochastic resonance, may underlie the extraordinary ability of some biological systems to detect and amplify small signals in noisy environments^{2–5}. Previous work has demonstrated the occurrence of stochastic resonance in a complex system of biological transducers and neural signal pathways⁶, but the possibility that it could occur at the sub-cellular level has remained open. Here we report the observation of stochastic resonance in a system of voltage-dependent ion channels formed by the peptide alamethicin. A hundred-fold increase in signal transduction induced by external noise is accompanied by a growth in the output signal-to-noise ratio. The system of ion channels considered here represents the simplest biological system yet known to exhibit stochastic resonance.

Voltage-sensitive ion channels in cell membranes are essential for primary biological signal transduction, playing important roles in the generation of nerve action potentials, synaptic transmission, and other critical cellular functions⁷. These channels switch randomly (in response to thermal fluctuations) between conductive and nonconductive states, but this switching is correlated with many external variables. These variables include external electric fields, the concentration of channel blocking or activating molecules and ionic strength. The possibility of a noise-induced improvement of weak signal detection in ion channels has been predicted^{1,3,8} but has not yet been observed. Studies intended to investigate the role of internal noise in ion channels⁹ and neuron behaviour¹⁰ have been inconclusive.

We have analysed the signal transduction properties of the alamethicin channel¹¹ in the absence and presence of an external electric noise source using the experiment schematically shown in Fig. 1. The polypeptide alamethicin (relative molecular mass ~2,000), promotes the formation of ion channels in lipid bilayers in a highly voltage-dependent manner. An average alamethicin-induced conductance increases by *e* times every 4 or 5 mV, depending on bilayer lipid composition¹². The molecular mechanism responsible for this voltage sensitivity is not fully established. One possible mechanism, shown in Fig. 1*a* inset, is the realignment of the dipoles of alamethicin helices in response to a change in the transmembrane potential. Application of an

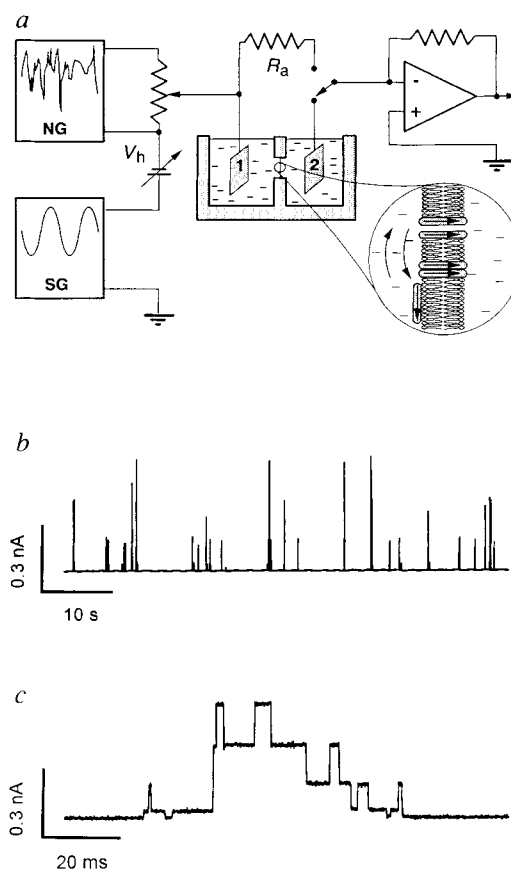


FIG. 1 Experimental apparatus (a) and output currents at a 5-mV (r.m.s.) 0.5-Hz sine-wave signal and zero external noise (b, c).

METHODS. A sine-wave signal from a Hewlett-Packard 3310B function generator (SG) was added to a d.c. holding potential of 130–150 mV and applied to electrode 1 through low-ohmic circuitry (output resistance <3 kΩ). Currents through the lipid bilayer with the incorporated voltage-dependent ion channels (inset), with total resistance always >10 MΩ, were recorded via the 'virtual ground' electrode 2 using an operational amplifier with 10 MΩ–1 GΩ feedback resistors. A passive resistor *R_a* had a value to the current through the match ion channels at a particular holding potential. A white noise generator (NG) of our own design was used to study the influence of external noise on the signal transduction in the system. The alamethicin channels were reconstituted in L-α-diphtanoyl lecithin bilayer membranes. Membrane-leakage d.c. current was <10⁻¹² A, membrane capacitance was ~50 pF. The steady-state conditions of measurement were achieved by equilibrating the bilayer and the aqueous salt solution (1 M NaCl) for 2–3 h after alamethicin addition. Other experimental details are as described previously¹⁴. The inset in a shows a lipid bilayer doped with alamethicin molecules. Arrows drawn on the 'molecules' represent molecular dipole moments responsible for the alamethicin channel voltage dependence^{11,12}. Molecules adsorbed on the membrane surface (molecules lying parallel to the bilayer) change their orientation when they interact with applied electric field. Channel formation thus involves the effective 'gating charge' derived from considering molecular dipole moment, number of molecules in the channel, and lipid bilayer thickness and dielectric constant. The transition between conductive (open) and non-conductive states (molecular clusters) of the channel is shown by reaction arrows.

electric field increases the number of molecules oriented across the membrane to facilitate channel opening. Ion channels are expressed in a stochastic manner as 'current bursts' rising from the background (Fig. 1*b, c*). Though seemingly random, the occurrences of current burst onsets in the presence of a 5-mV sine wave signal are correlated in time.

The power spectral density in Fig. 2*a* displays these correlations. There is a prominent peak at the applied signal frequency.

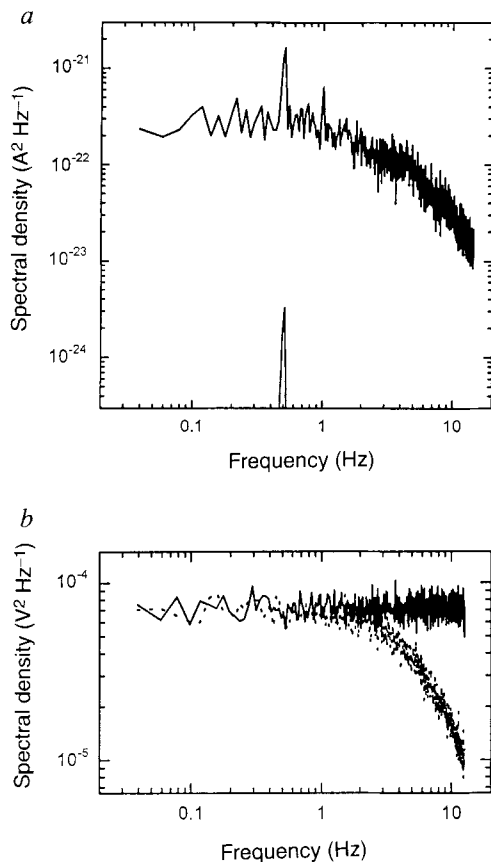


FIG. 2 *a*, Spectrum of the output current in the frequency range 0.02–20 Hz, showing correlations between moments of current burst onset. The signal component is much stronger than that expected from a passive linear circuit of the same average conductance (the peak at the bottom was measured for a $1.7 \times 10^{10} \Omega$ resistor). A 5 mV (r.m.s.) signal of 0.5 Hz was added to a 130 mV (d.c.) holding potential. No external noise was introduced at this stage. The spectrum is an average over a 20-min sample. *b*, Output voltage spectral density of our noise generator at the maximum output used in our experiments (solid line). Before application to the membrane, the output signal was filtered by an RC circuit to limit the noise spectral width to 5.3 Hz (dotted line), thus giving a value of 20 mV (r.m.s.). The basic element of the generator was a 1 G Ω resistor whose equilibrium Johnson noise was amplified. The signal was stationary, gaussian and with zero mean value (deviations from zero were within ± 0.1 mV at the maximum output and 30-min averaging).

The bottom spectral peak is measured using the circuit illustrated in Fig. 1*a* where the bilayer is replaced by a resistor, R_a , representing an average conductance of the system.

Ion channels produce a higher output signal than would a passive linear system of equal conductance, even including contributions from membrane capacitance. This gain can be characterized quantitatively by computing output power spectral density at the signal frequency, and subtracting the average noise background measured in the immediate vicinity of the signal peak (Fig. 2*a*). By formally introducing an empirical amplification coefficient α^2 as the ratio of signal intensities at the outputs of these two systems, we obtain a 27 dB gain for these particular experimental parameters.

The nature of this signal amplification can be understood if we consider a strong dependence of the probability of channel formation, $P(V)$, on the transmembrane potential V (ref. 12). For $P(V) \ll 1$ we have

$$P(V) \propto \exp(neV/kT) \quad (1)$$

where n is the effective 'gating charge' in units of elementary charge e . Condition $P(V) \ll 1$ holds for all our experiments. In particular, it is immediately seen for the recordings of ion current in Fig. 1*b*, where the combined life-time of the channel is much smaller than the record overall time.

If alamethicin channels are in 1- α -diphytanoyl lecithin bilayers, then n is close to 5. Now, taking into consideration the periodic nature of the signal, it is easy to show that for small signals in the absence of external noise the ratio can be expressed as

$$\alpha^2 = (1 + neV_h/kT)^2 \quad (2)$$

where V_h is a holding potential across the membrane, k is the Boltzmann constant, and T is the absolute temperature. We studied signal transduction properties of ion channels at different holding potentials, signal frequencies and signal amplitudes. Our measurements show that the gain increases monotonically with signal amplitude, does not depend appreciably on frequencies ranging from 0 to 2 Hz, and quickly decreases at higher frequencies. For slow small signals (2 or 3 mV r.m.s.), the empirical amplification coefficient approaches the theoretically predicted value. For example, for the particular conditions as specified in Fig. 2*a* legend, we expect α^2 to be $\sim 7 \times 10^2$ which is in a good agreement with the measured ratio.

To study the interaction between external noise and signal transduction, we measured the amplitude of the output signal, and the signal-to-noise ratio at the system output, as a function of external noise intensity. Our results (Fig. 3) show that the addition of external noise ranging from 0 to 20 mV (r.m.s.) to the input significantly increases the output signal (up to 35 dB), at an approximately constant signal-to-noise ratio (circles). Five-millivolt (r.m.s.) sine-wave signals at 0.2 Hz (filled symbols) and 0.5 Hz (open symbols) were used. It should be noted that at some intermediate values of noise intensities, a small increase in the signal-to-noise ratio is observed (Fig. 3 inset)—a characteristic feature of systems showing stochastic resonance¹. Small

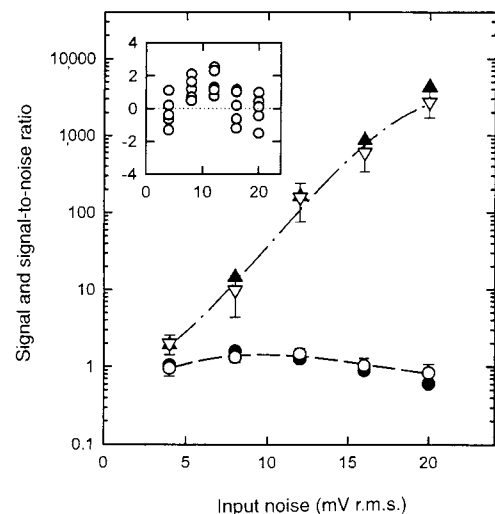


FIG. 3 A major noise-induced increase in signal output is achieved at an essentially constant, or even slightly increased, signal-to-noise ratio. The output signal (triangles) and the signal-to-noise ratio (circles) are given in units of their values at zero input noise. Sine waves of 5 mV (r.m.s.) and 0.2 Hz (filled symbols) and 0.5 Hz (open symbols) served as the input signal. The output signal-to-noise ratio was defined as the ratio of the signal peak maximum (Fig. 2*a*) to the background noise calculated as an average over spectral components in the immediate peak vicinity. Inset, the statistics of signal-to-noise ratio measurements on a finer scale over the same noise voltage range. The vertical axis is labelled in dB units. Each point represents an average over a 20-min signal recording.

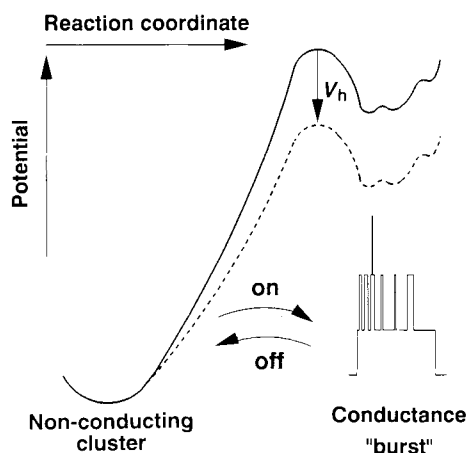


FIG. 4 The energy diagram of a quasi-bistable profile describing transitions of an alamethicin channel between non-conducting and conducting aggregates. The probability distribution along a reaction coordinate is sensitive to the transmembrane voltage mostly at the level of the transition between two main energy wells. In a conducting state corresponding to the higher-energy well, the channel voltage sensitivity is virtually lost. Transitions within this well describe flickering of the channel between different sub-levels during the conductance 'burst' shown in Fig. 1c. A small sine-wave signal modulates the probability of channel opening, introducing correlations in the moments of current burst (Figs 1 and 2). These correlations are the basis of the mechanism of electrical signal transduction through the system of voltage-dependent ion channels. Figure 3 shows that this transduction can be enhanced by application of random noise to the system input.

amplitudes of external noise increase both the signal and the noise at the system output, but the signal grows faster.

In our simple system, up to 100 parallel ion channels (formed by alamethicin) control the current by switching in a stochastic manner between the closed state and one of their open states. The probability of the channel being open strongly depends on the applied membrane voltage, whereas the relative probabilities of different conductive states are only weak functions of the voltage. A schematic diagram of the 'quasi-bistable' potential modulated by transmembrane voltage (Fig. 4) can explain such behaviour. The energy difference between the main potential wells is governed by the transmembrane voltage; the energy profile within 'burst' substates does not depend significantly on the voltage. A membrane with the alamethicin channel can therefore be viewed as a voltage-driven two-state dynamic system. A time-periodic signal applied to this system modulates the probability distribution between the two main states of the channel. Adding white noise to the signal increases the probability of conducting states. This happens at moments when the instantaneous value of the noise adds in phase with the positive half-waves of the applied signal.

Our results for a parallel array of voltage-dependent ion channels (representing a system of low biological complexity) suggest a novel mechanism of channel function regulation by a fluctuating environment. We have shown that addition of external noise to the input of this system induces a significant enhancement of signal transduction at constant (or, for some intermediate noise values, even increasing) signal-to-noise ratio at the output.

Biological signal processing often outperforms modern electronic devices⁴. Sensory transduction and neural signal processing are good examples of biological systems that function in highly fluctuating environments. In electronic instruments, the first amplification stage is designed to produce a high gain with minimal noise passed into the processing system. Special noise-rejection algorithms were invented that unfortunately always limited performance of signal processing devices. Perhaps nature devises amplification methods that can use an ambient noise to

improve signal transduction. Some of these amplifiers could be utilizing voltage-dependent ion channels in the way our system does.

Understanding signal processing and amplification by living organisms will require research at different levels of complexity. Our studies demonstrate that even the system of this simplicity (parallel voltage-dependent ion channels reconstituted in a lipid bilayer) is capable of ambient noise utilization. This is in accord with a recent model of stochastic resonance in parallel systems¹³ which showed that stochastic resonance might be a much more general phenomenon than previously thought and might apply to diverse systems not previously considered. □

Received 18 July; accepted 17 October 1995.

1. Wiesenfeld, K. & Moss, F. *Nature* **373**, 33–36 (1995).
2. Block, S. M. in *Sensory Transduction* 1–17 (Rockefeller Univ. Press, New York, 1992).
3. Polk, C. in *Proc. 13th A. Meeting Nat. Council Radiat. Protection & Measmt* (National Council on Radiation Protection, Washington DC, in the press).
4. Kalmijn, A. D. *Science* **218**, 916–918 (1982).
5. Bulsara, A. R., Lowen, S. B. & Rees, C. D. *Phys. Rev. E* **49**, 4989–5000 (1994).
6. Douglass, J. L., Wilkens, L., Pantazelou, E. & Moss, F. *Nature* **365**, 337–340 (1993).
7. Hille, B. *Ionic Channels of Excitable Membranes* (Sinauer Associates Inc., Sunderland, MA, 1992).
8. Makeyev, V. M. *Biophysika* **38**, 189–195 (1993).
9. Petracchi, D., Pellegrini, M., Pellegrino, M., Barbi, M. & Moss, F. *Biophys. J.* **66**, 1844–1852 (1994).
10. Pantazelou, E., Dames, C., Moss, F., Douglass, J. & Wilkens, L. *Int. J. Bifurc. Chaos* **5**, 101–108 (1995).
11. Hall, J. E., Vodyanov, I., Balasubramanian, T. M. & Marshall, G. R. *Biophys. J.* **45**, 233–247 (1984).
12. Eisenberg, M., Hall, J. E. & Mead, C. A. *J. Membrane Biol.* **14**, 143–176 (1973).
13. Collins, J. J., Chow, C. C. & Imhoff, T. T. *Nature* **376**, 236–238 (1995).
14. Bezrukov, S. M. & Vodyanov, I. *Biophys. J.* **64**, 16–25 (1993).

ACKNOWLEDGEMENTS. We thank V. Adrian Parsegian for valuable discussions. This work was supported by the US Office of Naval Research.

Effect of high salt concentrations on water structure

R. Leberman* & A. K. Soper†

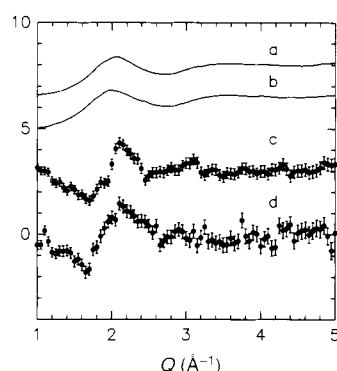
* European Molecular Biology Laboratory, Grenoble Outstation, BP 156, 38042 Grenoble Cedex, France

† ISIS Facility, Rutherford Appleton Laboratory, Chilton, Didcot, Oxfordshire OX11 0QX, UK

THE characteristic tetrahedral structure of water is known to be disrupted by changes in pressure and temperature^{1–3}. It has been suggested that ions in solution may have a similar perturbing effect^{4,5}. Here we use neutron diffraction to compare the effects of applied pressure and high salt concentrations on the hydrogen-bonded network of water. We find that the ions induce a change in structure equivalent to the application of high pressures, and that the size of the effect is ion-specific. Ionic concentrations of a few moles per litre have equivalent pressures that can exceed a thousand atmospheres. We propose that these changes may be understood in terms of the partial molar volume of the ions, relative to those of water molecules. The equivalent induced pressure of a particular ion species is correlated with its efficacy in precipitating, or salting-out, proteins from solution⁶.

Salting-out and salting-in are processes whereby concentrated aqueous salt solutions fractionate, concentrate and crystallize proteins and other biological macromolecules from solution. One of the first studies of the phenomenon of salting-out was performed by Hofmeister⁶, who showed that different salts have different efficiencies at salting-out egg-white protein, and that some salts do not cause salting-out at all. The phenomenon of salting-out or -in is not yet understood but the conventional view⁷ is that competition between dissolved salt and dissolved protein for water of hydration results in a loss or gain in solubility

FIG. 1 Hydrogen-hydrogen (HH) partial structure factors for 2 m ammonium sulphate (shifted upwards by 8; curve a) and 4 m ammonium chloride (shifted upwards by 6.5; curve b) as a function of neutron wavevector change, Q , from neutron diffraction data. The difference between these two functions is shown in curve c, multiplied by a factor of 10 and shifted upwards by 3. Note that the structural features in this difference do not coincide with the peaks in the HH partial structure factors for either solution, which rules out the possibility that the observed change is a simple scaling effect, and suggests there is a measurable change in the water structure between the two solutions. For comparison, the change in the HH partial structure factor for pure water per unit change in density is shown in curve d, scaled by a factor of 0.033.



of the protein. Hofmeister observed that at high concentrations sodium ions are more effective precipitants than ammonium ions, and that sulphate ions are more effective than chloride. We therefore looked at effects on water structure of these two pairs of ions.

In a typical 1:1 or 2:1 electrolyte in aqueous solution, there are ten partial structure factors to be determined (assuming compound ions like SO_4^{2-} are treated as a single component). This means that a single diffraction experiment cannot provide all the necessary information. Neutron diffraction using isotope substitution solves this problem by labelling specific sites within the solution isotopically and looking at the correlations associated with only those labelled sites. This approach has already taught us a great deal about the hydration water around ions in solution⁸, but much less is known about the structure of water itself in ionic solutions. Only one previous experiment has explored the structure of water in a concentrated ionic solution⁴, and this used hydrogen/deuterium isotope substitution of the water protons to extract the water-water (hydrogen-hydrogen, HH) radial distribution function. That experiment determined directly the degree of correlation between water molecules, and confirmed that water structure is modified quite radically at high concentrations of salt.

We have determined the water-water HH correlation functions for two pairs of solutions, namely ammonium chloride versus ammonium sulphate, and sodium chloride versus sodium sulphate. For each pair of solutions the concentration of cations (moles of ammonium or sodium per 1,000 g light water respectively) was kept the same, so that water structure was compared between each pair of solutions under similar charge-density conditions. The chosen concentrations were 2 molal for ammonium and sodium sulphates and 4 molal for ammonium and sodium chlorides. (In the case of the ammonium salts, the determined HH function is actually a composite function of water-water HH correlations, water-ammonium

HH correlations and ammonium-ammonium HH correlations, because of the exchange of hydrogen between water and the ammonium ion. But because the concentration of ammonium ions was the same for the two solutions, the changes in the ammonium-water and ammonium-ammonium correlations between the two solutions are not likely to be significant. Therefore on taking a difference between the HH structure factors for the two solutions, the ammonium-water and ammonium-ammonium contributions will, to a good approximation, cancel).

The data were recorded on the SANDALS diffractometer at the ISIS pulsed neutron source, because this instrument, with its large array of detectors at low scattering angles, is designed specifically for neutron measurements on samples containing low atomic masses. They were normalized to the scattering from a standard vanadium plate and corrected for container scattering, multiple scattering and attenuation using standard techniques⁹. They were then reduced to partial structure factors using existing methods¹⁰. The difference between the HH partial structure factors was calculated for the three cases $(\text{NH}_4)_2\text{SO}_4$ - NH_4Cl , Na_2SO_4 - NaCl , and NaCl -pure water. A fourth difference, NH_4Cl -pure water, was not useful because of the presence of ammonium-ammonium and ammonium-water correlations in the HH function, as described above.

Figure 1 shows the two HH partial structure factors associated with the ammonium ion and the difference between them. Also shown in this figure is the change in the HH function for pure water with change in pressure at constant temperature. For the latter curve the HH partial structure factor was determined for pure water (using identical apparatus to that described in ref. 2) at two pressures, namely 1 bar and 2.1 kbar, both at room temperature ($\sim 20^\circ\text{C}$). The higher pressure corresponds to an increase in water density of $\sim 7.2\%$. The two HH partial structure factors were differenced (high pressure minus low) and divided by the density change to give the change in the HH partial structure factor of water per unit density change (curve d in Fig. 1)¹¹.

All four differences, three for pairs of aqueous solutions and one for pure water, were then Fourier-transformed to real space and are shown in Fig. 2. All four difference functions have quali-

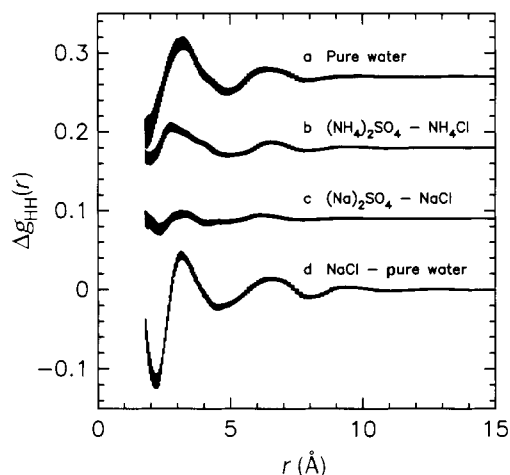


FIG. 2 Changes in the HH pair correlation function: curve a, for pure water for an applied density increase of 0.0017 molecules per $\text{\AA}^3$ above ambient density (0.0334 molecules per $\text{\AA}^3$); curve b, between a 2 m aqueous solution of $(\text{NH}_4)_2\text{SO}_4$ and a 4 m solution of NH_4Cl ; curve c, between a 2 m aqueous solution of Na_2SO_4 and a 4 m solution of NaCl ; and curve d, between a 4 m aqueous solution of NaCl and pure water. The qualitative similarity of the four curves is striking, and suggests that the primary effect of these salts on water structure is equivalent to a significant density increase in the water around the component ions compared to pure water under ambient conditions.

TABLE 1 Estimated equivalent pressures of some aqueous solutions

Solution pair	Equivalent water density change (molecules per $\text{\AA}^3$)	Equivalent pressure change (kbar)
NaCl -pure water	0.0017	1.4
Na_2SO_4 - NaCl	0.0003	0.3
$(\text{NH}_4)_2\text{SO}_4$ - NH_4Cl	0.0009	0.8

The density change in the water is estimated from the data in Fig. 2, by comparing the amplitude of the oscillations in curves b, c, d with those for pure water (curve a). To obtain the equivalent pressure, it is assumed that the density of water changes by 0.0012 molecules per $\text{\AA}^3$ for a 1-kbar pressure change¹⁴.

tatively the same shape, that is, a negative region near radius $r=2$ Å, a positive region near $r=3.2$ Å, and a set of oscillations with a period of ~ 3 Å for radius values beyond this. Under ambient conditions the intermolecular HH pair correlation function for pure water consists of two main peaks, one at ~ 2.4 Å and the other at ~ 3.8 Å with a minimum between them at $r \approx 3.0$ Å (ref. 12). These two peaks are caused by the strong orientational correlations between neighbouring water molecules in hydrogen-bonded water. The fact that in ionic solutions and with the application of pressure to pure water the 2.4-Å and 3.8-Å peaks are apparently diminished, while the minimum at 3.0 Å fills in, implies that the orientational correlations become more disordered as pressure is applied to pure water. Exactly the same arguments are used to explain the lowering of the melting point of water as pressure is applied.

Our data on the differences of the HH radial distribution for different pairs of dissolved salts (Fig. 2) suggests that a similar kind of disordering mechanism must occur in aqueous solutions of ionic salts, but the size of the effect depends significantly on the particular pair of ions involved. The effect is especially pronounced for the difference NaCl–pure water. In every case the change in HH function is as if the pressure of the water has been increased. Using the data on pure water as a calibration, it is possible to estimate roughly the corresponding equivalent pressure difference for each of these pairs of combinations; the values are given in Table 1.

Note that the observed changes could also be regarded as equivalent to those that occur when the temperature of pure water is increased at constant density as orientational correlations become more disordered in both cases². But for a temperature increase at constant water density, there is in any case a corresponding (large) pressure increase. Because the aqueous solution measurements described here have been carried out at room temperature, we believe that it is more accurate to characterize the structural change as due to a density or pressure change in the water at constant temperature, rather than by an equivalent temperature change of the solvent at constant density. Sodium chloride has a large equivalent pressure (1.4 kbar; Table 1), the equivalent pressure of sodium sulphate is marginally higher still, $(1.4 + 0.3 = 1.7$ kbar; Table 1—note that this additive value agrees well with the direct measurement) whereas ammonium sulphate causes a significantly higher equivalent pressure than ammonium chloride. It is interesting that these trends fit qualitatively with the Hofmeister series, as both Na_2SO_4 and $(\text{NH}_4)_2\text{SO}_4$ salt-out readily, whereas NaCl is less effective than Na_2SO_4 at precipitating proteins, and NH_4Cl is not effective at all⁶.

To some extent, these equivalent pressures can be understood by considering the partial molar volumes of the respective salts. Both ammonium and chloride ions have partial molar volumes close to that of water ($\sim 18 \text{ cm}^3 \text{ mol}^{-1}$)¹³, and these values do not change appreciably with concentration¹⁴. On the other hand, both sodium and sulphate ions have significantly smaller partial molar volumes than water which suggests that these ions have a substantial electrostrictive effect on water structure, and so introduce disorder into the normal tetrahedral coordination of water molecules. □

Received 21 September; accepted 17 October 1995.

1. Postorino, P., Tromp, R. H., Ricci, M. A., Soper, A. K. & Nielson, G. W. *Nature* **366**, 668–670 (1993).
2. Postorino, P., Ricci, M. A. & Soper, A. K. *J. chem. Phys.* **101**, 4123–4132 (1994).
3. Tromp, R. H., Postorino, P., Neilson, G. W., Ricci, M. A. & Soper, A. K. *J. chem. Phys.* **101**, 6210–6215 (1994).
4. Tromp, R. H., Neilson, G. W. & Soper, A. K. *J. chem. Phys.* **96**, 8460–8469 (1992).
5. Enderby, J. E., Howells, W. S. & Howe, R. A. *Chem. Phys. Lett.* **21**, 109–112 (1973).
6. Hofmeister, F. *Arch. Exp. Path. Pharmacol.* **24**, 247–260 (1888).
7. Zubay, G. *Biochemistry* 3rd Edn (Brown, Dubuque, IA, 1993).
8. Herdman, G. J. & Neilson, G. W. *J. molec. Liq.* **46**, 165–179 (1990).
9. Soper, A. K., Howells, W. S. & Hannon, A. C. *ATLAS: Analysis of Time-of-flight Diffraction Data from Liquid and Amorphous Samples* (Rep. No. 89-046, Rutherford Appleton Lab., Chilton, 1989).
10. Soper, A. K. & Luzar, A. *J. chem. Phys.* **97**, 1320–1331 (1993).

11. Soper, A. K. & Turner, J. Z. *Int. J. mod. Phys.* **7**, 3049–3076 (1993).
12. Soper, A. K. & Phillips, M. C. *Chem. Phys.* **107**, 47–60 (1993).
13. Millero, F. J. *Chem. Rev.* **72**, 147–176 (1971).
14. *Handbook of Chemistry and Physics* 55th Edn D194–D236 (CRC, Cleveland, 1974).

ACKNOWLEDGEMENTS. Part of this work was performed under an EC Twinning Agreement.

Liquid-crystalline phases as templates for the synthesis of mesoporous silica

George S. Attard, Joanna C. Glyde & Christine G. Göltner

Department of Chemistry, University of Southampton, Southampton SO17 1BJ, UK

THE synthesis of inorganic mesoporous materials using ionic surfactant template molecules was first reported in 1992^{1,2}, and surfactant-mediated synthesis has since been used to form a variety of mesoporous materials^{3–8}. Such materials could find application in catalysis, membrane and separation technology, and molecular engineering. Previous syntheses have used low surfactant concentrations, and the templating mechanism (which is still controversial) is thought to be a cooperative process involving the interaction of inorganic ions with discrete surfactant aggregates^{1,2,5,9}. Here we report the templating of silica mesostructures from ordered liquid-crystalline mesophases: the resulting silica phase, with pores of ~ 3 nm diameter, is a cast of the organic mesophase. As the phase diagram of the surfactant/silica/water system at high surfactant concentration is similar to the (known) phase diagram of surfactant/water alone, this approach should introduce an element of predictability into the synthesis of mesoporous materials.

In our experiments we employed the liquid-crystalline phases of the non-ionic surfactants octaethylene glycol monododecyl ether (C_{12}EO_8) and octaethylene glycol monohexadecyl ether (C_{16}EO_8) to template the synthesis of mesoporous silica. The liquid-crystalline behaviour of these surfactants has been studied in detail¹⁰. Tetramethyl orthosilicate (TMOS) was employed as the ceramic precursor as its hydrolysis and condensation behaviour under a range of temperature and pH conditions has been investigated extensively^{11,12}.

Type I (that is, normal topology) hexagonal phases (H_1) were prepared by mixing each surfactant with water acidified to pH 2

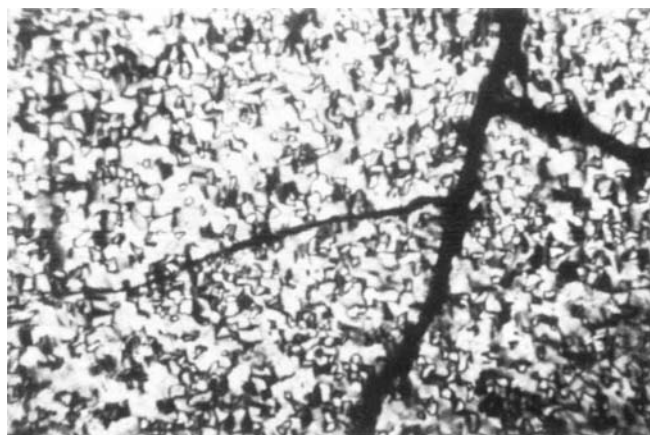


FIG. 1 The formation of silica in the hexagonal mesophase of C_{12}EO_8 . Photomicrograph of solid birefringent material at 130°C ($\times 65$) showing cracks due to shrinkage. Mixture preparation: $\text{C}_{12}\text{EO}_8\text{:H}_2\text{O(pH 2):TMOS} = 1:1:2$ (by weight); methanol removed by placing mixture in gentle vacuum for 15 min.

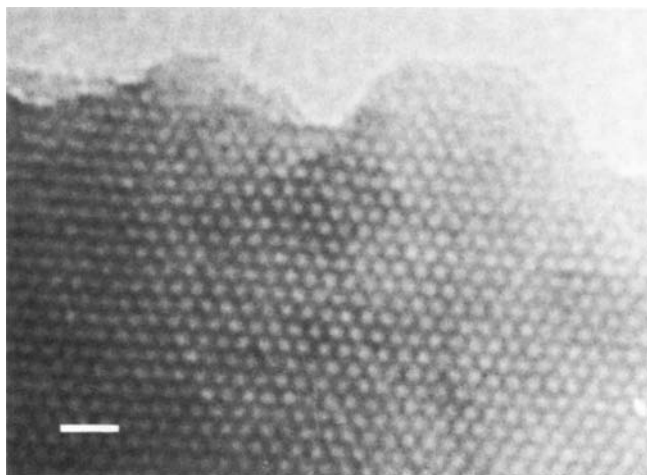


FIG. 2 Transmission electron micrograph showing the ordered hexagonal structure of silica synthesised in the H_1 phase of $C_{12}EO_8$ and calcined at 450 °C (12 h under N_2 , 16 h under O_2). Scale bar, 10 nm.

(using hydrochloric, sulphuric or nitric acid). In both preparations the ratio of surfactant to water was 50 wt%. TMOS was added to these mixtures such that the final amount of TMOS was not more than 0.25 mole equivalents with respect to the water. Thus in these mixtures the water content is equal to, or higher than, that required for the stoichiometric hydrolysis of the silica precursor. Methanol is released from the hydrolysis of the TMOS and destroys the H_1 phase, leaving a viscous isotropic liquid. However, when the methanol is removed using gentle vacuum the H_1 phase is reformed. When viewed under a polarizing microscope these mixtures appear birefringent and mobile and consist of a single phase. The optical textures are identical to those of the H_1 phases formed in the corresponding surfactant/water mixtures. Bulk mixtures were kept at room tempera-

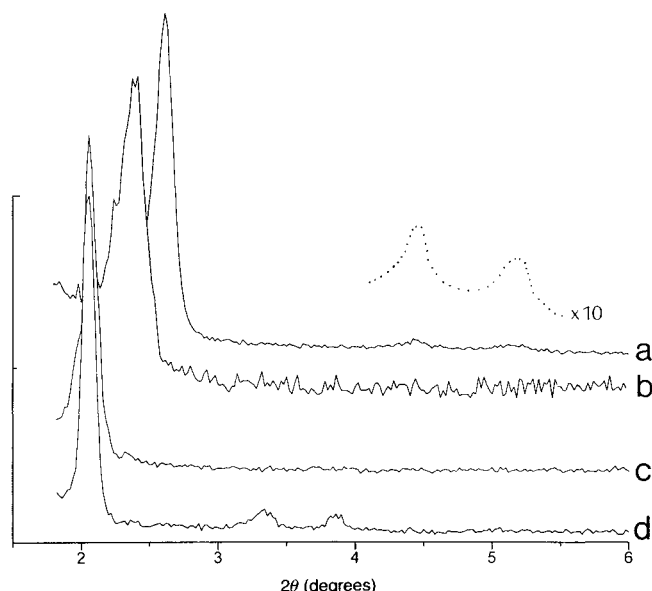


FIG. 3 a–d, X-ray diffraction patterns. Trace a, calcined silica synthesized in the H_1 phase of $C_{12}EO_8$; trace d, the hexagonal phase of $C_{12}EO_8 + H_2O$ (50 wt%). Diffractograms b and c were obtained at different times following the removal of methanol from the ternary system $C_{12}EO_8/H_2O/TMOS$. Trace b, solid sample (12 h after the start of the reaction); trace c, fluid sample (15 min after the start of the reaction).

ture for 18 hours. During this time the fluid mixtures became progressively more viscous, and eventually set into rigid monoliths that had the shape of the container. The formation of the silica network in a thin film sample was monitored using polarized light microscopy. To increase the rate of condensation, the fluid birefringent samples were heated to 45 °C and held at this temperature for several hours. As in the case of the bulk samples, the thin film preparations became progressively more viscous, becoming solid after 2–3 hours. The optical textures of the materials remained unaltered during the progress of the reaction. Heating the solidified samples at 5 °C min⁻¹ to above 100 °C resulted in the appearance of small cracks (Fig. 1). The optical textures remained unchanged even when the organic matter within the preparations decomposed on further heating (~350 °C). The hexagonal mesostructures of the aged surfactant/water/silica mixtures were confirmed by transmission electron microscopy (TEM) and, in the case of the $C_{12}EO_8$ preparation, also by X-ray diffraction. In the absence of silica precursor the H_1 phase of $C_{12}EO_8$ melts to form a micellar solution at 58 °C.

The ordered structures that remained after the removal of organic matter by calcination were investigated by TEM and X-ray diffraction. TEM studies of the silica obtained from the H_1 phase of $C_{12}EO_8$ showed an hexagonal array of regularly sized

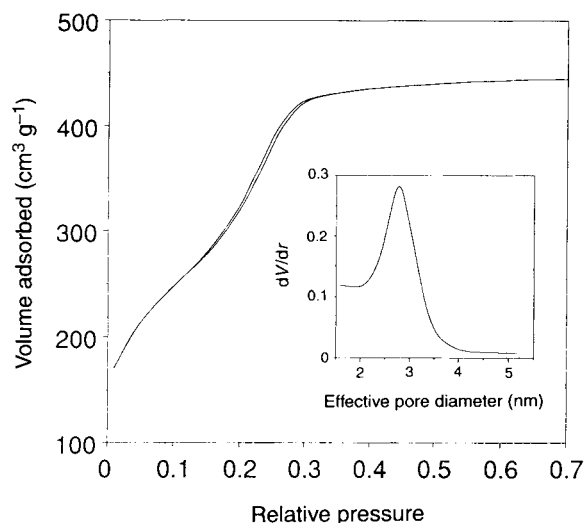


FIG. 4 Nitrogen adsorption-desorption isotherms for silica synthesized in the H_1 phase of $C_{12}EO_8$. (Relative pressure is P/P_0 , where P is the equilibrium pressure of the adsorbate and P_0 is the saturation pressure of the adsorbate at the temperature of the adsorbent. Volume adsorbed is at STP). Inset, the corresponding Horvath-Kawazoe plot for the adsorption data (dV/dR is the derivative of the nitrogen volume adsorbed with respect to the pore diameter of the adsorbent).

holes of 28 Å diameter (Fig. 2) separated by ~12 Å-thick silica walls. X-ray diffraction patterns obtained from the calcined silica (Fig. 3a) show a d_{100} reflection at $2\theta = 2.62^\circ$, which corresponds to a periodicity of 34 Å. From this, the hole-to-hole distance is estimated to be 39 Å, which is consistent with the value of 40 Å obtained from TEM studies. X-ray diffraction patterns were obtained at various times during the course of the reaction. The results show that the system initially has a lattice parameter of 43 Å (Fig. 3c), identical to that measured for the hexagonal phase of the surfactant/water mixture (Fig. 3d). As the mixture solidifies, the lattice parameter decreases to 38 Å. The specific surface area of the calcined silica was found to be 1,400 (±10) m² g⁻¹ from nitrogen BET adsorption measurements. This is comparable to values reported for mesoporous ceramics synthesized via other routes^{1,2,4,5,7,8}. Nitrogen adsorption/desorption curves (Fig. 4) showed virtually no hysteresis, indicating a highly uniform pore size with no pore-blocking effects from narrow

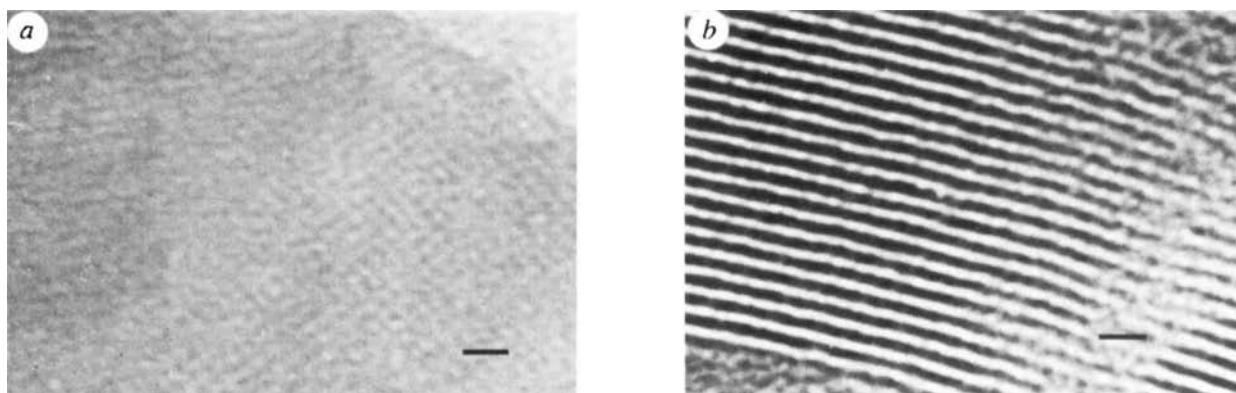


FIG. 5 Transmission electron micrographs of calcined silica synthesised in the Ia3d cubic phase (a), and in the L_a phase (b) of $C_{16}EO_8$. Scale bars, 10 nm.

pores during desorption. The pore size distribution curve that is obtained from the desorption data peaks at 2.7 nm, and has a half-height width of 0.8 nm (Fig. 4 inset). These results are in excellent agreement with the data obtained from TEM studies.

To investigate whether mesophases other than the hexagonal phase could be used as templates, we prepared samples of $C_{16}EO_8$ /TMOS/water that were either in the Ia3d cubic phase or in the lamellar phase. We chose $C_{16}EO_8$ rather than $C_{12}EO_8$ because the former has cubic and lamellar phases that are stable over wider composition ranges. As in the case of the hexagonal preparations, single-phase systems were observed throughout the condensation process (18 h at 30 °C). The monolithic solid obtained from the cubic phase was optically isotropic, whereas the solid from the lamellar phase had an optical texture characteristic of this phase. TEM micrographs of the solids obtained after calcination (Fig. 5a, b) show that cubic and lamellar mesophases can template the formation of the corresponding mesostructures.

To obtain a direct comparison between our approach and previous syntheses, we set up TMOS/water mixtures with low concentrations (~1 wt%) of the non-ionic surfactants. It should be noted that these concentrations are higher than the critical micellar concentrations of the two surfactants. These mixtures produced fine precipitates which were found to have heterogeneous mesostructures upon investigation by TEM. These findings are in agreement with the results described by Pinnavaia¹³. In a further set of experiments, we used the cationic surfactant cetyl trimethyl-ammonium bromide (CTAB) instead of the non-ionic surfactants. Mixtures of CTAB/silica precursor/water at low pH and low surfactant concentration (~1 wt%) produce mesoporous solid precipitates as reported previously^{4,5}. At higher CTAB/water ratios (~50 wt%) hexagonal mesophases were obtained. Condensation of the silica precursor in these systems did not disrupt the phase, and monolithic birefringent solids were obtained after 18 h at 35 °C. Calcination of the samples and subsequent TEM studies showed them to have hexagonal mesostructures. Furthermore, the process is independent of the acid used to adjust the initial pH of the mixture (that is, sulphuric, nitric or hydrochloric acid).

In the case of low pH and low surfactant concentration, three factors have been proposed as being crucial for successful templating: (1) multidentate binding of positively charged silicate oligomers to the surfactant molecules via an anion (electrical triple layer); (2) polymerization of inorganic species at the interface, and (3) charge-density matching at the interface⁵. The fact that continuous silica networks develop throughout the aqueous domains of the liquid crystalline phases of non-ionic surfactants and CTAB suggests that in the high surfactant concentration regime the formation of mesostructured composites is effectively independent of factors (1) to (3) above. Furthermore, the

absence of regular mesostructures in materials prepared with low concentrations of non-ionic surfactants suggests that it is the liquid-crystalline phase, and not micellar aggregates, that directs the formation of the mesoporous materials. □

Received 5 December 1994; accepted 17 October 1995.

1. Kresge, C. T., Leonowicz, M. E., Roth, W. J., Vartuli, J. C. & Beck, J. S. *Nature* **359**, 710–712 (1992).
2. Beck, J. S. *et al.* *J. Am. chem. Soc.* **114**, 10834–10843 (1992).
3. Inagaki, S., Fukushima, Y. & Kuroda, K. *J. chem. Soc., chem. Commun.* 680–682 (1993).
4. Behrens, P. & Stucky, G. D. *Angew. Chem., int. Edn. engl.* **32**, 696–699 (1993).
5. Monnier, A. *et al.* *Science* **261**, 1299–1303 (1993).
6. Behrens, P. *Adv. Mater.* **5**, 127–132 (1993).
7. Huo, Q. *et al.* *Nature* **368**, 317–321 (1994).
8. Alfreðsson, V. *et al.* *J. chem. Soc., chem. Commun.* 921–922 (1994).
9. Cheng, C.-F., Luan, Z. & Klinowski, J. *Langmuir* **11**, 2815–2819 (1995).
10. Mitchell, D. J., Tiddy, G. J. T., Waring, L., Bostock, T. & McDonald, M. P. *J. chem. Soc., Faraday Trans. I* **79**, 975–1000 (1983).
11. Chen, K. C., Tsuchiya, T. & Mackenzie, J. D. *J. non-cryst. Solids* **81**, 227–237 (1986).
12. Mackenzie, J. D. *J. non-cryst. Solids* **100**, 162–168 (1988).
13. Bagshaw, S. A., Prouzet, E. & Pinnavaia, T. J. *Science* **269**, 1242–1244 (1995).

ACKNOWLEDGEMENTS. We thank B. Cressey and T. Khan for assistance with the TEM studies, D. Currie for assistance with the X-ray measurements, and M. Weiler for the use of his X-ray diffractometer. This work was supported by the UK Engineering and Physical Sciences Research Council.

Constraints on the origins of hydrocarbon gas from compositions of gases at their site of origin

Leigh C. Price* & Martin Schoell†

* US Geological Survey, MS-940, Denver Federal Center, Denver, Colorado 80225, USA

† Chevron Petroleum Technology Co., PO Box 446, La Habra, California 90633-0446, USA

It is widely accepted that natural gas is formed from thermal decomposition of both oil in reservoirs and, to a lesser extent, the organic matter in shales from which the oil was derived^{1–6}. But laboratory pyrolysis experiments on shales do not reproduce the methane-rich composition typical of most gas reservoirs⁷, leading to suggestions⁷ that other mechanisms, such as transition-metal catalysis, may be important. The discrepancy might, however, instead arise because gas (and oil) deposits have migrated from their source rocks, so that the reservoir composition might not be representative of the composition in the source rocks where the hydrocarbons were generated. To address this question, we have

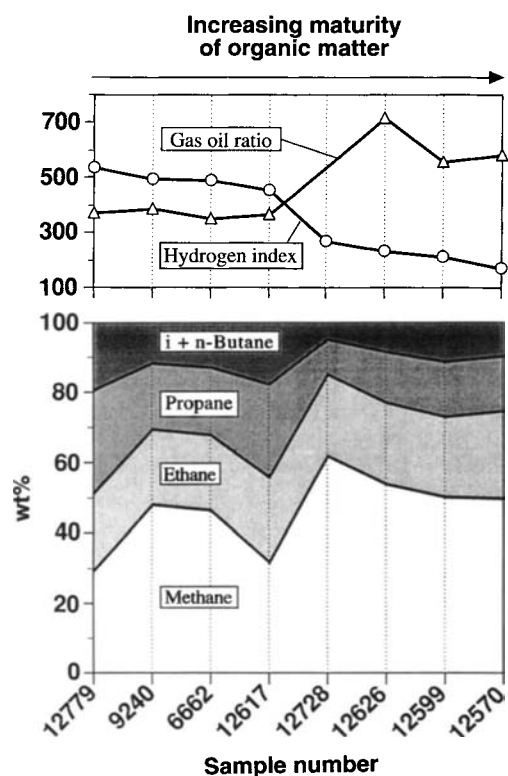


FIG. 1 Normalized weight per cent (wt%) of gases from Bakken oils. Maturity, as gauged by decreasing Rock-Eval hydrogen indices, increases left to right. Bakken shale gas-to-oil ratios (in standard cubic feet of gas per barrel of oil) are also plotted for each sample. The vertical scale to the left of the plot is the index for both parameters. Note that shale hydrogen indices strongly decrease with increasing maturity whereas gas-to-oil ratios strongly increase, reflecting the increasing amount of gas cogenerated with oil with increasing maturity. The 'sample number' on the horizontal scale of the bottom plot is the North Dakota Geological Survey well number (complete data on wells and oils are available in ref. 11). Rock-Eval is a petroleum-geochemical analytical instrument which measures a rock's capacity for oil generation and its maturity, the hydrogen index reflecting both properties. High hydrogen indices characterize immature rocks (low burial temperatures) and low hydrogen indices characterize mature rocks (higher burial temperatures). Hydrogen indices in Fig. 1 vary from immature (oil generation just beginning, NDGS-12779) to mature (oil generation largely complete, NDGS-12570), thus reflecting almost the full range of oil generation possible in nature. Vitrinite reflectance (R_o) is the most common petroleum-geochemical maturity measurement, measuring the percentage of light reflected from plant (vitrinite particles recovered from shales) which increases with maturity. Vitrinite reflectance values are unavailable here because of their strong suppression in the hydrogen-rich organic matter of the Bakken shales²⁰. Data from Table 1.

analysed gas samples coproduced with oils directly from a source rock (the Bakken shales, North Dakota, USA) where the local geology has prevented significant hydrocarbon migration. The methane contents of these Bakken-shale gases are much lower than that of conventional gas reservoirs, but are consistent with that from pyrolysis experiments^{8,9} on these shales. Thus, because these Bakken gases form with (rather than from) oils, we argue that compositional differences between gases from source rocks and conventional gas deposits result from fractionation processes occurring after hydrocarbon expulsion from the source rock.

Several different sources of thermogenic natural gases are generally recognized¹⁻⁶. The most important sources are thought to be thermal decomposition of oil in reservoirs, and/or organic matter in shale, to gas. Burial temperatures higher than those involved with the generation of oil from source rocks are thought to be involved (150–200 °C, and above). Another possible source

of gas is cogeneration with oil in source rocks during initial hydrocarbon generation (burial temperatures of 100–150 °C). Most thermogenic gas is generally thought to originate from the higher-temperature processes, with gases cogenerated with oil in source rocks making up a much smaller percentage of gas deposits¹⁰. Moreover, gases generated at high burial temperatures are thought to be methane-rich (80 mol%, 70 wt%, and usually greater), with the gases becoming even more methane-rich as temperature increases⁶. However, because gases are very mobile in the subsurface, obtaining unaltered gas samples from their place of origin is usually not possible. Thus, these models are not easily tested.

Numerous wells in the Williston basin in North Dakota have produced economically valuable oils (containing dissolved natural gas) from the two shales of the Upper Devonian and Lower Mississippian Bakken Formation and from directly adjacent rocks. These oils are unconventional because they have never been expelled from the immediate area of their source rock, instead residing in fractures in otherwise non-porous (non-reservoir) rocks. In contrast, 'conventional oils' are usually expelled from the immediate area of their source rock, and undergo some ('secondary') migration to form deposits in the indigenous pore space ('porosity') of conventional reservoir rocks, such as limestones or sandstones. Here, conventional oils may be mixed with natural gas generated from another source.

Analyses of the unconventional Bakken oils and of conventional oils from the principal oil reservoir of the same basin (the limestone units in the middle of the Lower and Upper Mississippian Madison Group) demonstrate that these oils are two distinct and genetically unrelated families¹¹. Moreover, neither mixing of Bakken and Madison oils within a single reservoir nor the occurrence of Bakken oil in a conventional reservoir on the US side of the basin have been encountered¹¹. These facts and other analyses¹² demonstrate that the Bakken shales and adjacent rocks form a tight petroleum system, wherein Bakken-generated hydrocarbons remained in, or near, their source rock—a natural 'closed' pyrolysis system which locally trapped all generated products. Our analyses (Fig. 1) of eight gases,

TABLE 1 Composition of gases coproduced with Bakken oils

Sample number	HI	GOR	C ₁	C ₂	C ₃	ISO-C ₄	n-C ₄
12779	535	373	29.38	22.96	28.42	4.70	14.54
			52.00	20.32	18.49	2.29	7.09
9240	495	388	47.60	22.50	18.44	2.94	8.52
			69.24	16.39	9.77	1.18	3.42
6662	491	350	45.42	22.13	19.98	3.06	9.41
			67.94	16.55	10.37	1.26	3.88
12617	453	366	31.59	24.89	26.16	4.51	12.85
			54.15	21.34	16.31	2.13	6.07
12728	268	ND	61.01	24.43	9.84	1.36	3.37
			78.10	15.64	4.58	0.48	1.19
12626	234	717	58.03	24.49	14.13	2.43	5.92
			72.78	16.81	7.05	0.92	2.24
12599	212	556	49.63	23.89	15.20	2.91	8.38
			70.68	17.01	7.87	1.14	3.29
12570	170	584	48.99	26.22	15.19	2.55	7.04
			69.71	18.66	7.86	1.00	2.76

Normalized weight per cent (first line), and volume per cent (second line) of C₁–C₄ hydrocarbon gases coproduced with Bakken oils. North Dakota Geological Survey well (sample) numbers are in the first column. Rock-Eval hydrogen indices (HI) are in the second column, and decrease down the table (maturity increases downwards). Gas-to-oil ratios (GOR) in standard cubic feet of gas per barrel of oil are in the third column. C₁ is methane, C₂ ethane, C₃ propane, iso-C₄ iso-butane, and n-C₄ normal-butane. Gas caps do not exist in these unconventional Bakken reservoirs. Oil gravities range from 0.785 to 0.825 g cm⁻³, and much other data on the oils are available¹¹. Gas samples were obtained from separators (oil and gas producing equipment) using stainless-steel cylinders and flowing gas through the cylinder, shutting first the exit valve then the entrance valve. All gas samples were from the areally limited 'fairway' of active commercial horizontal Bakken oil drilling near the Bakken southwestern depositional edge, North Dakota. Changes in the character of the shale organic matter (facies variations) are thus insignificant.

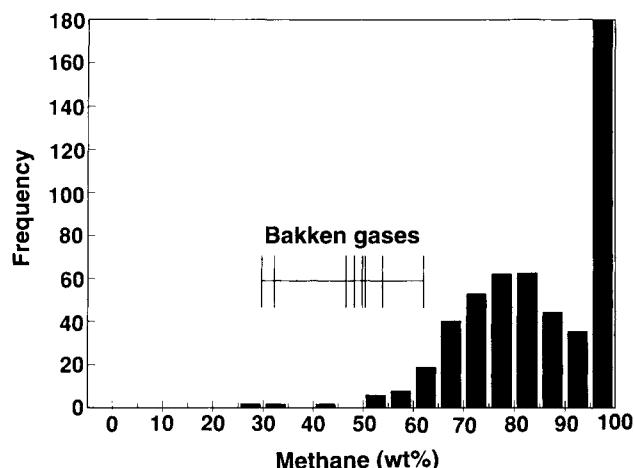


FIG. 2 Weight per cent of methanes from the eight Bakken gases of this study (the vertical lines) superimposed on Fig. 1 from Mango *et al.*⁷ which shows 500 analyses of natural gas deposits randomly selected by Mango *et al.*⁷ from ref. 18. Note that the Bakken gases have much lower contents of methane (much higher C_2 – C_4 contents) than almost all the analysed gases from gas deposits.

coproduced with these oils from their source rock, provide significant insight into gas compositions at their generation site and thus also into origins of natural gas. For example, gas-to-oil ratios for all Bakken oils produced in the basin range from 200 to 4,000, and average ~1,100 standard cubic feet of gas per barrel of oil. Thus, the very existence of these gases and their elevated gas-to-oil ratios demonstrate that substantial gas is cogenerated with oil in source rocks long before thermal destruction of the larger C_{15+} hydrocarbons occurs.

Bakken shale samples were also obtained either from the same wells as the gas samples or from wells no further than 1.6 km from the producing well. Rock-Eval hydrogen indices¹¹ of these shales (Fig. 1) are highly variable, revealing that the individual shales (and gases) were exposed to much different burial temperatures in the geological past (that is, they have different 'maturities').

As expected from existing models, the Bakken gases become more methane-rich with increasing maturity (Fig. 1), although this trend is irregular with several reversals. But even the most mature gases have far greater amounts of C_2 – C_4 components than usual for conventional gas deposits (Fig. 2).

Extensive analyses of conventional gas deposits demonstrate that natural gases are methane-rich^{7,13}, often exceeding 90 wt% methane (C_1) normalized to C_1 – C_4 hydrocarbon gases (Fig. 2). However, gases generated from oil source rocks in laboratory experiments in closed, water-wet systems have much higher C_2 – C_4 contents. Thus, Mango and co-workers⁷ argued that these pyrolysis experiments were not replicating gas generation in nature, and proposed a catalytic pathway involving transition metals in the source rock for hydrocarbon gas generation in nature to explain this discrepancy.

Note that the gases from Bakken oils have 39–71 wt% C_2 – C_4 components (Fig. 1) compared with the ≤ 20 wt% (Fig. 2) typical of deep-basin 'dry-gas' deposits or gas caps over deep-basin oil deposits. The high C_2 – C_4 percentages in Figs 1 and 2 are not expected from current models. Moreover, high percentages of C_2 – C_4 components (43–69 wt%) were produced in aqueous-pyrolysis experiments on the Bakken shale^{8,9}. Bakken shale organic matter is typical of marine-derived oil-source rocks, which have sourced most of the world's discovered oil¹⁰. Thus, it seems unlikely that the elevated C_2 – C_4 components in gases from Bakken oils, and from experiments on Bakken shales, are from an unusual organic matter. Rather, the data suggest that the

large gas volumes cogenerated with oil in source rocks are more generally rich in C_2 – C_4 gases than are gases in deposits. Thus we conclude that fractionation of hydrocarbon gases occurs after they are expelled from source rocks to account for the compositional differences between gases in source rocks versus gases in deposits. This post-expulsion fractionation could occur either along the 'secondary migration' path towards a reservoir or in the oil or gas reservoir itself.

Such a 'migration-fractionation' of gases has been hypothesized and attributed⁹ to both preferential condensation¹⁴ of the C_2 – C_4 gases in oil and buoyancy-driven processes. In this model, with further migration of methane-rich late-stage gas into the oil reservoir (with a gas cap), the reservoir becomes completely filled. Because of the buoyancy differences between oil and gas (gas resides above oil in the reservoir), oil with a high C_2 – C_4 content is displaced from the reservoir by Gussow's principle of differential entrapment¹⁵. These displaced fluids undergo secondary migration to shallower reservoirs in the basin, with the original reservoir eventually becoming completely filled with a methane-rich gas. Thus, we believe that 'migration-fractionation', besides maturity and mixing^{1,2,16,17}, may also affect the final composition of gas deposits. If the high C_2 – C_4 gas contents of both the produced and remaining-in-place oils in a basin were mass-balanced with the composition of the gas deposits in that basin, then the entire system would have a much higher C_2 – C_4 gas content than previously portrayed.

Another possible cause, albeit a relatively minor one, for the different compositions of the gases from Bakken oils and those in large sample bases^{13,18} is a sampling error. As mentioned above, the C_2 – C_4 gases are known to preferentially condense into (and be coproduced with) oil¹⁴. Yet some samples from large reported gas-sample bases^{13,18} are from wells producing gas caps over depleted oil deposits, and such samples do not represent the gas composition of the entire system (gas cap plus the C_2 – C_4 gases in the oil).

Thus the reported high methane percentages in many gas deposits may result largely from a fractionation of the C_1 – C_4 gases after expulsion from source rocks. Transition-metal catalysis during gas generation⁷ does not seem necessary to explain the methane-rich nature of gas deposits, a compositional feature which apparently does not necessarily result from natural gas generation in source rocks. We also note that the Bakken shales have high concentrations of transition metals, similar to the rock studied by Mango *et al.*⁷, and might therefore be taken as ideally suited to transition-metal catalysis. But whereas Mango *et al.*⁷ observed gases with 80–87 wt% methane in their experiments (favouring transition-metal catalysis), Bakken gases from producing wells and laboratory pyrolysis experiments have only 29–61 and 31–57 wt% methane, respectively. Thus the Bakken gas data suggest that if transition-metal catalysis is indeed playing a role in the generation of gas from source rocks, it might not be a significant factor in determining gas compositions. We note that McNeil and BeMent¹⁹ have attributed the results observed in the experiments of Mango *et al.*⁷ to a reaction pathway other than transition-metal catalysis, and have concluded that transition-metal catalysis may have not been operable in those experiments at all. We feel it is probable that the reaction pathway McNeil and BeMent¹⁹ propose plays a significant role in hydrocarbon gas generation.

Mass-balance calculations suggest that the Bakken shales have generated at least 300 billion barrels of oil⁹, which essentially remains in the shales, and more so in the adjacent rocks, constituting a very large in-place oil resource¹¹. Gas-to-oil ratios of produced Bakken oils average about 1,100 standard cubic feet of gas per barrel of oil, implying an in-place gas-resource of 330 trillion cubic feet of C_2 – C_4 -rich gas. The data of this study demonstrate that C_2 – C_4 -rich gas is cogenerated, in abundance, with oil and also suggest that such cogenerated gas may provide a significant contribution to gas deposits, more generally even dry-gas deposits. This possibility, as noted⁹, implies that much

larger in-place gas resources may exist than those previously extrapolated from the model of gas originating solely from oil destruction, because much more gas is cogenerated with oil in source rocks than would ever be available from thermal destruction of oil deposits. □

Received 4 January; accepted 9 October 1995.

1. Schoell, M. *Bull. Am. Ass. Petrol. Geol.* **67**, 2225–2238 (1983).
2. James, A. T. *Bull. Am. Ass. Petrol. Geol.* **67**, 1176–1191 (1983).
3. Schoell, M. *Chem. Geol.* **71**, 1–10 (1988).
4. Gailimov, E. M. *Chem. Geol.* **71**, 77–96 (1988).
5. Chung, H. M., Gormally, J. R. & Squires, R. M. *Chem. Geol.* **71**, 97–104 (1988).
6. Clayton, C. *Mar. Petrol. Geol.* **8**, 232–240 (1991).
7. Mango, F. D., Hightower, J. W. & James, A. T. *Nature* **368**, 536–538 (1994).
8. Wenger, L. M. & Price, L. C. in *Organic Geochemistry—Advances and Applications in Energy and the Natural Environment* (ed. Manning, D. A. C.) 335–339 (European Association Organic Geochemists, Manchester, UK, 1991).

9. Price, L. C. in *Sources of Natural Gas* (eds Rice, D. & Schoell, M.) *Chem. Geol.* (in the press).
10. Tissot, B. P. & Welte, D. H. *Petroleum Formation and Occurrence* (Springer, New York, 1984).
11. Price, L. C. & LeFever, J. *Bull. Can. Petrol. Geol.* **42**, 187–218 (1994).
12. Price, L. C. & LeFever, J. in *Geological Studies Relevant to Horizontal Drilling: Examples from Western North America* (eds Schmocker, J. W., Coalson, E. B. & Brown, C. A.) 199–214 (Rocky Mountain Association Geologists, Denver, 1992).
13. Rice, D. D., Threlkeld, C. N. & Vuletic, A. K. *Open-File Rep.* 88–391 (US Geol. Surv., 1988).
14. Salisbury, G. P. *Mem. Am. Ass. Petrol. Geol. Mem.* 9, V. 2 1433–1445 (1968).
15. Gussow, W. C. *Bull. Am. Ass. Petrol. Geol.* **38**, 816–853 (1954).
16. Jenden, P. D., Newell, K. D., Kaplan, I. R. & Watney, W. L. *Chem. Geol.* **71**, 117–147 (1988).
17. Fuex, A. N. J. *Geochem. Explor.* **7**, 155–188 (1977).
18. Namak, J. E. & Sigler, S. *Information Circ.* No. 9301 (Bureau of Mines, US Dept of Interior, Pittsburgh, 1991).
19. McNeill, R. I. & BeMent, W. O. *Energy and Fuels* (in the press).
20. Price, L. C. & Barker, C. E. J. *Petrol. Geol.* **8**, 59–84 (1985).

ACKNOWLEDGEMENTS. We appreciate reviews by A. James, M. Lewan, P. Jenden, S. Larter, F. Mango and especially E. Shock. We thank Meridian Oil Incorporated for allowing collection of gas samples from their producing wells, and J. LeFever for her help with sample collection.

Radiocarbon evidence for extensive plate-boundary rupture about 300 years ago at the Cascadia subduction zone

Alan R. Nelson^{*†}, Brian F. Atwater[‡],
Peter T. Bobrowsky[§], Lee-Ann Bradley^{*},
John J. Clague^{||}, Gary A. Carver[¶],
Mark E. Darienzo^{*}, Wendy C. Grant[‡],
Harold W. Krueger[☆], Rodger Sparks^{**},
Thomas W. Stafford[†] Jr & Minze Stuiver^{††}

^{*} US Geological Survey, MS 966, Box 25046, Denver, Colorado 80225, USA

[†] Institute of Arctic and Alpine Research, CB 450, University of Colorado, Boulder, Colorado 80309-0450, USA

[‡] US Geological Survey at Department of Geological Sciences, Box 351310, University of Washington, Seattle, Washington 98195-1310, USA

[§] British Columbia Geological Survey Branch, 1810 Blanshard Street, Victoria, British Columbia V8V 1X4, Canada

^{||} Geological Survey of Canada, 100 West Pender Street, Vancouver, British Columbia V6B 1R8, Canada

[¶] Department of Geology, Humboldt State University, Arcata, California 95521, USA

[#] Geology Department, Portland State University, Box 751, Portland, Oregon 97207, USA

[☆] Krueger Enterprises, Inc., Geochron Laboratories Division, Cambridge, Massachusetts 02138, USA

^{**} Rafter Radiocarbon Laboratory, Nuclear Sciences Group, Institute of Geological and Nuclear Sciences, Ltd, Box 31 312, Lower Hutt, New Zealand

^{††} Department of Geological Sciences and Quaternary Research Center, Box 351310, University of Washington, Seattle, Washington 98195-1310, USA

THE Cascadia subduction zone, a region of converging tectonic plates along the Pacific coast of North America, has a geological history of very large plate-boundary earthquakes^{1,2}, but no such earthquakes have struck this region since Euro-American settlement about 150 years ago. Geophysical estimates of the moment magnitudes (M_w) of the largest such earthquakes range from 8 (ref. 3) to 9½ (ref. 4). Radiocarbon dating of earthquake-killed vegetation can set upper bounds on earthquake size by constraining the length of plate boundary that ruptured in individual earthquakes. Such dating has shown that the most recent rupture, or series of ruptures, extended at least 55 km along the Washington coast within a period of a few decades about 300 years ago⁵. Here we report 85 new ¹⁴C ages, which suggest that this most recent

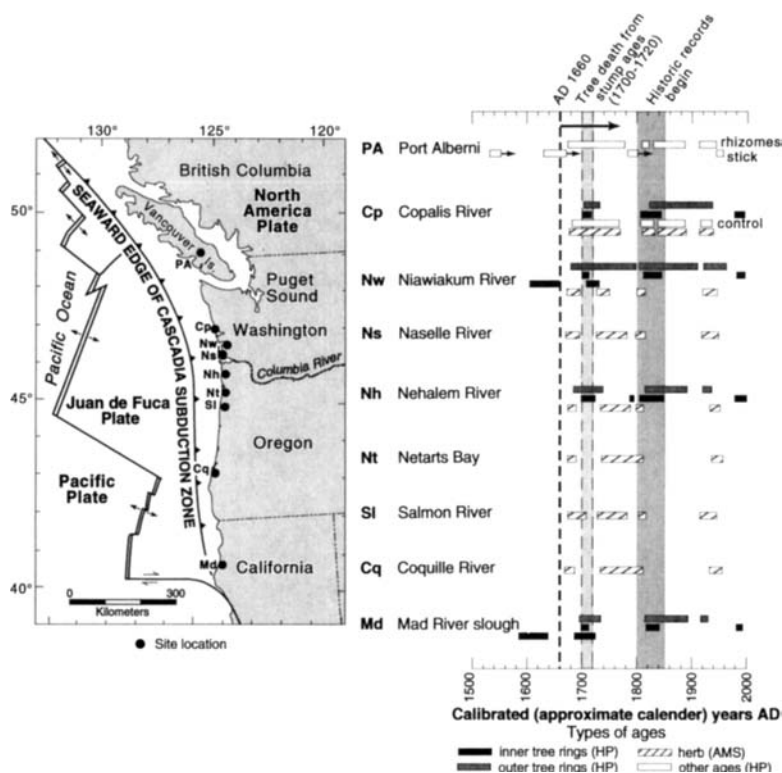
rupture (or series) extended at least 900 km between southern British Columbia and northern California. By comparing the ¹⁴C ages with written records of the past 150 years, we conclude that a single magnitude 9 earthquake, or a series of lesser earthquakes, ruptured most of the length of the Cascadia subduction zone between the late 1600s and early 1800s, and probably in the early 1700s.

Earthquake size increases with rupture length. Whereas ruptures rarely extend more than 250 km during subduction-zone earthquakes near M_w 8 (refs 6–8), ruptures nearly 1,000 km long produced the largest twentieth-century earthquakes— M_w 9.2 in Alaska and M_w 9.5 in Chile⁹. We infer possible magnitudes of plate-boundary earthquakes along the Cascadia subduction zone by estimating the length of fault rupture during prehistoric earthquakes. Although the Cascadia plate boundary is not exposed on land, rupture length can be estimated from the extent of sudden coseismic subsidence along the subduction zone. Such subsidence, inferred to record elastic extension of the North America plate or faulting within it^{3,10,11}, caused sudden submergence and burial of wetland soils at many estuaries along the Pacific coast from northern California to southern British Columbia^{1,11–17}. Entombed tree stumps and herbaceous plants rooted in the tops of buried soils^{18,19}, commonly capped by sheets of tsunami-deposited sand^{14,16,20–22}, reinforce the inference of earthquake-induced subsidence for the youngest of the buried soils at many Cascadia estuaries.

We used two methods of radiocarbon dating and several types of plant fossils to date the time that the youngest soil suddenly subsided (Fig. 1). High-precision ¹⁴C ages measured in Seattle were obtained by proportional counting of gas derived from blocks of tree rings from the outer and inner parts of stumps of trees killed by submergence at four sites (Table 1; ages from sites Cp and Nw on Fig. 1 were reported earlier⁵). Other high-precision ages at our most northerly site (PA, Fig. 1) came from a stick in tsunami-laid sand above the youngest soil and from below-ground stems (rhizomes) and attached leaf bases of a tidal-marsh herb that probably grew in the sand soon after its deposition. For each of three of the stump sites (Cp, Nw and Nh) and four other sites (Fig. 1), we used accelerator mass spectrometer (AMS) ¹⁴C methods at the New Zealand laboratory to date the stems and leaves of seven or eight individual herbs (Table 1). These herbs probably died, like the trees, from the effects of earthquake-induced subsidence^{18,19}.

These procedures increase the resolution of the dating in two ways. (1) Dating of the earthquake-killed plants reduces uncertainty in relating the age of dated material to the time of an earthquake. Annual rings in the trees show how many years probably elapsed between formation of the dated rings and the time of the earthquake(s) that killed the trees¹⁹. The AMS-dated herbs probably grew within a few years of the subsidence that killed them¹⁸. However, dated parts of many of the herbs include

FIG. 1 Cascadia subduction zone, sites sampled for ^{14}C dating, and calendar age ranges corresponding to single ^{14}C ages and means of ^{14}C ages from each site (Table 1). Horizontal bars in column on right show calendar age ranges at two standard deviations calculated from ^{14}C ages (at one standard deviation) using error multipliers of 1.0 (New Zealand laboratory) and 1.6 (Seattle laboratory)^{29,31}. Blank, shaded and pattern fills within bars mark ranges for Port Alberni (PA) high-precision (HP) ages and means of other types of ages. Small arrows on bars for the stick age from PA indicate that it is a maximum age for tsunami deposition. Ranges for stump ages have been shifted to the right by 5 years (outer-ring ages), 40 years (most inner-ring ages), 60 years (lower inner-ring range at Md), or 80 years (lower inner-ring range at Nw); such shifts are illustrated in Fig. 2 for ages from a single stump at Md. Herb and outer-ring age ranges indicate that subsidence occurred after AD 1660 (heavy dashed vertical line); historical records of the past 200 years restrict subsidence to before about 1850 and probably before about 1800 (vertical shaded bar). Shaded interval between 1700 and 1720 shows likely time of subsidence inferred from the least ambiguous ages from southern Washington and northern California.



perennial stems that may predate subsidence by more than a few years (see below). (2) Analytical errors can be reduced by averaging ages (Table 1) from single plants that died at or about the same time²³. Contemporary death has been tested at Mad River slough (Md, Fig. 1), where matching the ring-width patterns implies that the dated stumps died within four years of one another¹⁹. Ages grouped by sample type at each site passed a χ^2 test for contemporaneity²³ at the 95% level.

Mean ages for each sample type (Table 1) do not differ statistically among sites. Stump means overlap at one standard deviation, both for inner rings and for outer rings. Although many herb means differ from each other at one standard deviation, all herb means passed the χ^2 test for contemporaneity at the 95% level.

Despite agreement by sample type, AMS ages on earthquake-killed herbs appear systematically greater than high-precision ages on earthquake-killed trees (Table 1). The mean age for eight herbs from the Nehalem tree (179 ± 15 ^{14}C -yr BP) is significantly greater than the mean age of rings averaging five years before tree death from three stumps at this site (128 ± 09 ^{14}C -yr BP). Herb ages are also greater, though not significantly at the 95% level, than a high-precision age on the outer 10 rings of a stump at the Copalis site. Further, the mean herb age of 161 ± 15 ^{14}C -yr BP at the Niawiakum River exceeds the range of ^{14}C ages that correspond (according to the calibration curve of Fig. 2) to AD 1700–1720, the likely time of tree death at this site (see below).

Neither of two possible explanations—systematic age differences between measurements from different laboratories, and dating of long-lived parts of herbs—appears to account fully for the disagreements between AMS-herb ages and stump ages. No systematic differences have been apparent between the Seattle and New Zealand laboratories in age comparisons among ^{14}C laboratories^{24,25}. In our only interlaboratory comparison, the outermost ring of a stump at the Copalis River ('control' in Table 1 and Fig. 1) gave a mean AMS age 26 ± 24 ^{14}C -yr greater than a high-precision age on the outer 10 rings of the stump. We speculate that some fraction of the woody or decay-resistant parts of many of the AMS-herb samples grew as much as one

or two calendar decades before the earthquake(s) that killed the nearby trees. Woody perennial stem bases made up 50–90% by weight of most samples of the herb *Potentilla pacifica* (70% of herb samples). Samples of the stems of *Juncus balticus* (15% of herb samples) included parts of its resistant rhizomes. However, dating of samples consisting mostly of the annual leaves of a sedge, probably *Carex lyngbyei*, at the Coquille River gave the greatest of the AMS mean ages (Table 1), and we found no relation between the proportion of woody or resistant material in dated samples and their ^{14}C ages.

Converted to ranges (95% confidence interval) of calibrated (approximately calendar) years, the means of the herb and outer-ring ages show that the most recent coseismic subsidence at each site occurred after AD 1660 (Fig. 1). This is the earliest date consistent with the herb means at each of the seven herb-dated sites. The outer-ring stump means exclude ages earlier than 1680 at three of the herb sites (Cp, Nw, Nh) and at Mad River slough.

Inner-ring ages from stumps narrow the time of subsidence in California and Washington to the early 1700s. Ages on three different blocks of rings from a stump at Mad River slough limit tree death to about 1700–20; these ages eliminate other possible dates by matching distinctive steps (wiggles) in the calibration curve that relates radiocarbon time to calendar time (Fig. 2). Similar wiggle matching for a stump at the Niawiakum River in southern Washington implies tree death between about 1710 and 1730 (refs 5, 26). If all the dated stumps in southern Washington died within a few years of one another, 40-year-before-death ages (inner-ring) from six of these stumps can be averaged and the resulting mean— 209 ± 06 ^{14}C -yr BP—can be used to further narrow the time of subsidence in southern Washington⁵. This time is between 1703 and 1715, as estimated with recently revised data for ^{14}C calibration²⁶.

Both high-precision ages from the tsunami deposit at Port Alberni are consistent with tsunami deposition after the middle 1600s (Fig. 1). The rhizomes, which should shortly postdate the tsunami deposit into which they grew, formed no earlier than 1670. The tsunami probably coincided with the most recent coseismic subsidence on the southwest coast of Vancouver Island²¹.

TABLE 1 Radiocarbon ages on plant remains near the top of the youngest buried soil

Site	Distance (km)*	No of ages	Least age	Weighted mean age	Greatest age
High-precision ages on inner tree rings that average 40 years before tree death					
Copalis River	740	3	199 ± 15	207 ± 09	219 ± 19
Niawiakum River	685	3	189 ± 20	210 ± 09	219 ± 13
Nehalem River	585	1		211 ± 13	
Mad River slough	60	2	202 ± 11	214 ± 08	225 ± 10
High-precision ages on inner tree rings that average 60 or 80 years before tree death					
Niawiakum River	685	1		301 ± 10	(rings 75–86)
Mad River slough	60	1		282 ± 14	(rings 59–61)
High-precision ages on outer tree rings that average five years before tree death					
Copalis River	740	1		112 ± 11	
Nehalem River	585	3	124 ± 14	128 ± 09	132 ± 11
Mad River slough	60	4	098 ± 14	120 ± 08	137 ± 14
High-precision age on outer tree rings 1–20					
Niawiakum River	685	1		152 ± 30	
High-precision ages on rhizomes and attached leaf bases of a tidal-marsh herb and a stick in tsunami-deposited sand					
Port Alberni	980	1		145 ± 12	(herb)
Port Alberni	980	1		261 ± 14	(stick)
AMS age on a control sample consisting of an outer tree ring formed one year before tree death					
Copalis River	740	7	091 ± 41	138 ± 17	177 ± 46
AMS ages on the rooted leaves and stems of tidal-marsh herbs					
Copalis River	740	7	067 ± 42	139 ± 17	165 ± 44
Niawiakum River	685	8	120 ± 41	161 ± 15	272 ± 47
Naselle River	660	8	121 ± 44	162 ± 16	232 ± 45
Nehalem River	585	8	102 ± 41	179 ± 15	275 ± 44
Netarts Bay	550	8	133 ± 63	188 ± 18	212 ± 41
Salmon River	510	7	114 ± 43	157 ± 17	199 ± 49
Coquille River	320	8	069 ± 71	192 ± 17	242 ± 45

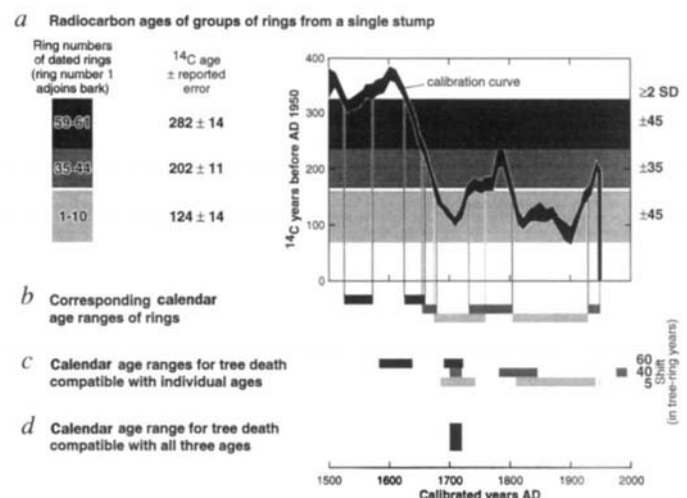
High-precision gas-proportional ^{14}C ages (in ^{14}C -years before AD 1950) were measured through long decay counting of large samples (original weight 60–90 g) in heavily shielded counters at the Seattle laboratory^{26,29}. Reported errors do not include an error multiplier estimated at ≤ 1.6 . Ages on inner and outer tree rings come from stumps of Sitka spruce (*Picea sitchensis*) with the exception of one outer-ring age measured on hardwood at the Nehalem River. The number of dated rings is 10 or 20 for Washington samples and 3–10 for Oregon and California samples. The two high-precision ages from Port Alberni were measured on the rhizomes (below-ground stems) and attached leaf bases of the herb *Triglochin maritima* and on a bark-free stick (species unidentified). A radon correction²⁶ accounts for the difference between some calendar age ranges reported here and in 1991⁵. AMS ages (in ^{14}C -years before AD 1950) were measured by accelerator mass spectrometry at the New Zealand laboratory. An estimate of total analytical error is reported with each age. AMS herb samples (5–62 mg) consist of leaves of Lyngby's sedge (*Carex* cf. *C. lyngbyei*) at Coquille River, and leaf bases and subaerial stems of Pacific silverweed (*Potentilla pacifica*) and soft rush (*Juncus* cf. *J. balticus*) at Copalis River, Niawiakum River, Naselle River, Nehalem River, Netarts Bay and Salmon River. Herb samples were cleaned before AMS analysis by scraping in water under a microscope and were pretreated three times using standard procedures³⁰. Two AMS accelerator targets were prepared for each herb sample from the same CO_2 gas. Except for 8 of 61 samples, where only one target yielded an age, the mean of the two target ages is the reported age for each of the other samples. Mean inner-ring and outer-ring stump ages and mean AMS-herb ages for each site were calculated through averaging weighted by the inverse square of laboratory-reported errors³¹. The standard deviation of the mean of eight AMS ages from the New Zealand laboratory on a stump control sample from Copalis River (138 ± 17 ^{14}C -yr BP; 'control' on Fig. 1) suggests that reported errors on AMS ages encompass all internal analytical variability^{25,32}.

* Distance north of Cape Medocino, which is at the southern end of the Cascadia subduction zone (Fig. 1).

Historical documents imply that the subsidence at all dated sites predates initial Euro-American exploration (latest AD 1700s) or settlement (early 1800s). The documents make no mention of earthquakes that could have caused coastal subsidence along a large part of the Cascadia subduction zone, particularly

after widespread Euro-American settlement of coastal areas began about 1850 (refs 27, 28). Such evidence against nineteenth- and twentieth-century subsidence eliminates many younger calendar age ranges not already ruled out by wiggle matching of stump ages (Figs 1 and 2).

FIG. 2 High-precision ^{14}C ages and corresponding calendar age ranges for selected tree rings cut from a single stump rooted in the youngest buried soil at Mad River slough (Md, Fig. 1). *a*, Ages for three different groups of rings intersect different parts of the ^{14}C calibration curve, which relates ^{14}C time to calendar time²⁶. Two-standard-deviation errors for each age, calculated using an error multiplier of 1.6, are shown on the right edge of the diagram. *b*, Horizontal bars show the corresponding calendar age ranges, shaded as in *a*. *c*, Ranges have been shifted to the right by the number of years between the dated rings and the bark. The time at which the dated tree could have died is limited to the narrow range shown in *d* (AD 1700–1720) where the ranges for all three ages overlap.



Taken together, the radiocarbon and historical evidence imply rupture along at least 900 km of the plate boundary (Table 1, Fig. 1), either in a single M_w 9 earthquake or through a series of lesser earthquakes. Because of uncertainties that include poorly understood disagreements between stump and herb ages (discussed above), the evidence merely limits the duration of possible series. Serial earthquakes could not have spanned more than the 190 years between AD 1660 and 1850. A series might have begun with a rupture of a few hundred kilometres off the Oregon coast, where we obtained the greatest of the mean herb ages (Cq, Nt and Nh, Fig. 1). But the occurrence of such a series is not confirmed by the least herb mean from Oregon (S1) or by the similarity between stump ages at the Nehalem River and those in southern Washington and northern California (Md). If a prolonged series of earthquakes began in California and proceeded

northwards (or began in Washington and went southwards), it could not have lasted as long as the series of three earthquakes near M_w 8 off Colombia and Ecuador between 1942 and 1979. These earthquakes ruptured, from south to north, about 500 km of subduction zone in 37 years⁷. By contrast, the earthquake or earthquakes that killed the dated trees at sites 680 km apart in Washington and California occurred in less than 20 years.

Having failed to show age differences as large as those in the Colombia-Ecuador example, our dating strengthens the possibility that one plate-boundary earthquake ruptured most of the length of the Cascadia subduction zone in the early 1700s. Such a rupture would have had an area larger than 50,000 km², as judged from rupture widths inferred from geodetic and heat-flow data¹⁰. This area implies an earthquake magnitude close to 9—near the upper end of the range of magnitudes considered possible at the Cascadia subduction zone^{3,4}. □

Received 14 June; accepted 23 October 1995.

- Atwater, B. F. et al. *Earthquake Spectra* **11**, 1–18 (1995).
- Nelson, A. R., Shennan, I. & Long, A. J. *J. geophys. Res.* (in the press).
- McCaffrey, R. & Goldfinger, C. *Science* **267**, 856–859 (1995).
- Heaton, T. H. & Hartzell, S. H. *Science* **236**, 162–168 (1987).
- Atwater, B. F., Stuiver, M. & Yamaguchi, D. K. *Nature* **353**, 156–158 (1991).
- Ando, M. *Tectonophysics* **27**, 119–140 (1975).
- Kanamori, H. & McNally, K. C. *Bull. seism. Soc. Am.* **72**, 1241–1253 (1982).
- Thatcher, W. J. *geophys. Res.* **95** (B3), 2609–2624 (1990).
- Plafker, G. J. *geophys. Res.* **77**, 901–925 (1972).
- Dragert, H., Hyndman, R. D., Rogers, G. C. & Wang, K. J. *geophys. Res.* **99**, 653–668 (1994).
- Clarke, S. H. Jr & Carver, G. A. *Science* **255**, 188–192 (1992).
- Peterson, C. D. & Darienzo, M. E. in *Earthquake Hazards in the Pacific Northwest of the United States* (eds Rogers, A. M., Walsh, T. J., Kockelman, W. J. & Priest, G. R.) (Open-File Rep. 91–441C, US Geological Surv., Denver, CO, 1991).
- Nelson, A. R. in *Quaternary Coasts of the United States—Marine and Lacustrine Systems* (eds Fletcher, C. P. & Wehmiller, J. F.) 287–301 (Spec. Publ. 48, Soc. for Sedimentary Geology, Tulsa, OK, 1992).
- Atwater, B. F. *J. geophys. Res.* **97** (B2), 1901–1919 (1992).
- Mathewes, R. W. & Clague, J. J. *Science* **264**, 688–691 (1994).
- Darienzo, M. E., Peterson, C. D. & Clough, C. J. *coastal Res.* **10**, 850–876 (1994).

- Nelson, A. R., Jennings, A. E. & Kashima, K. *Geol. Soc. Am. Bull.* (in the press).
- Atwater, B. F. & Yamaguchi, D. K. *Geology* **19**, 706–709 (1991).
- Jacoby, G., Carver, G. & Wagner, W. *Geology* **23**, 77–80 (1995).
- Clague, J. J. & Bobrowsky, P. T. *Quat. Res.* **41**, 176–184 (1994).
- Clague, J. J. & Bobrowsky, P. T. *Geol. Soc. Am. Bull.* **106**, 1293–1303 (1994).
- Hemphill-Haley, E. *Geol. Soc. Am. Bull.* **107**, 367–378 (1995).
- Ward, G. K. & Wilson, S. R. *Archaeometry* **20**, 19–34 (1978).
- Stuiver, M. & Pearson, G. W. in *Radiocarbon after Four Decades* (eds Taylor, R. E., Long, A. & Kra, R. S.) 19–33 (Springer, New York, 1992).
- Scott, E. M., Cook, G. T., Harkness, D. D., Miller, B. F. & Baxter, M. S. *Radiocarbon* **34**, 520–527 (1992).
- Stuiver, M. & Becker, B. *Radiocarbon* **35**, 35–65 (1993).
- Douthitt, N. *South Coast History* (River West, Coos Bay, OR, 1985).
- Rhonda, J. P. *Astoria and Empire* (Univ. of Nebraska Press, Lincoln, NE, 1990).
- Stuiver, M. & Pearson, G. W. *Radiocarbon* **28**, 805–838 (1986).
- Bradley, L. A. & Stafford, T. W. Jr *Radiocarbon* **36**, 399–405 (1994).
- Stuiver, M. & Reimer, P. J. *Radiocarbon* **35**, 215–230 (1993).
- Long, A. *Radiocarbon* **32**, 393–397 (1990).

ACKNOWLEDGEMENTS. We thank B. E. Benson for cutting blocks of rings, and P. J. Wilkinson for high-precision ¹⁴C analyses. Reviews by H. M. Kelsey, R. F. Madole, G. S. Burr and G. C. Rogers improved the manuscript. This work was supported by the National Earthquake Hazards Reduction Program of the US Geological Survey, the US Nuclear Regulatory Commission, the Geological Survey of Canada, and the British Columbia Geological Survey.

Quartz dissolution by the sponge *Chondrosia reniformis* (Porifera, Demospongiae)

Giorgio Bavestrello*, Attilio Arillo*, Umberto Benatti†, Carlo Cerrano*, Riccardo Cattaneo-Vietti*, Luciano Cortesogno‡, Laura Gaggero‡, Marco Giovine‡, Michela Tonetti† & Michele Sarà*

* Istituto di Zoologia dell'Università di Genova, via Balbi 5, I-16126 Genova, Italy

† Istituto di Chimica Biologica dell'Università di Genova, Viale Benedetto XV, I-16132 Genova, Italy

‡ Dipartimento di Scienze della Terra dell'Università di Genova, Sezione Mineralogia e Petrografia, Corso Europa 26, I-16132 Genova, Italy

MANY marine organisms etch calcareous substrata¹. Indeed sponges, mainly of the genus *Cliona*, are important factors in the erosion of calcareous coasts^{2,3}. Among terrestrial organisms, only a few lichens are known to penetrate siliceous rocks^{4,5}, an ability unknown in the animal kingdom. The Demospongiae have a siliceous skeleton formed by spicules⁶ of various shapes and sizes, but several species also incorporate sand grains or foreign spicules^{7,8}. The demosponge *Chondrosia reniformis* Nardo has no autochthonous spicules but incorporates a wide range of foreign materials in its ectosome^{9,10}. Here we report that quartz particles are strongly etched and made uniform in size, quickly and with sharp selectivity, the hydrated silica (chalcedony and opal) remaining unaltered. The

presence of a thick collagenous ectosome¹¹ suggests that ascorbic acid, the reducing agent in proline hydroxylation, might be involved in quartz etching by *C. reniformis*.

Observations of the ectosome of *Chondrosia reniformis* by a Philips 515 scanning electronmicroscope (SEM) equipped with a spectrometer in the energy-dispersive mode showed the presence of crystalline quartz, opaline sponge spicules, gypsum and several kinds of silicates, such as albite, muscovite, sanidine, chlorite and sodium amphibole. Although the silicates showed well shaped crystalline features or angular fracture surfaces, and the opaline spicules were well preserved (Fig. 1a), quartz grains always showed spiny corrosion patterns (Fig. 1a).

The sponge etching action was experimentally checked by using the following siliceous materials: marine sand (acid purified, grain diameter 125–250 µm, from Laigueglia, Ligurian Sea, Italy, composed of angular and poorly sorted quartz grains, mostly formed by monocrystalline fragments (Fig. 1b); laboratory sand (BDH) (comprising round grains of diameter 250–450 µm), the internal texture being composed of aggregates of polygonal granoblastic quartz (Fig. 1g); monocrystalline, prismatic, biterminate quartz grains (2 mm in size) (Fig. 1f); chalcedony; and opal. Each material was laid down on the ectosome of ten sponges, and grain size and shape within their ectosomes were checked by daily tissue corings for eight days. Engulfed chalcedony and opal showed perfectly conserved crushing surfaces after eight days, whereas monocrystalline prismatic biterminate quartz grains and both marine and BDH quartz showed a sharp decrease in grain size. Within a couple of days their pristine shape was obscured, and the size of the particles became uniform after eight days (Fig. 2).

SEM observations (Fig. 1) on incorporated marine (Fig. 1b–e) and BDH (Fig. 1g–h) sands showed partial or total etching of the particle surface. Differences in the pitting pattern between

marine and BDH sands were due to crystal composition. Owing to their large size, some monocrystalline, prismatic, biterminate quartz grains, which had settled on the sponge surface, were partly incorporated and etched (Fig. 1b).

To study the dissolution activity, various specimens of *Chondrosia reniformis*, were separately reared in artificial silica-free sea water (ASW) (see Fig. 3 legend for details). A BDH sand layer was laid down on the ectosome of a group of sponges, whereas untreated specimens and sand dispersed in ASW were used as blanks. An increase in the amount of dissolved silica in the water of the sand-treated sponges was evident 13 h after the beginning of the experiment (Fig. 3), but no increase was detected in the blank samples. Over the next 30 h, a sharp increase in the amount of dissolved silica in the medium with the sandy sponges was recorded, reaching an average value of $0.104 \pm 0.005 \text{ mg l}^{-1}$.

Some polyvalent organic acids, notably citric and oxalic acids, have been shown to dissolve quartz¹². The presence in *Chondro-*

sia of a thick ectosome composed of collagen, and the evidence that this anatomical region is the centre of quartz dissolution, led us to test for a possible involvement in quartz corrosion of ascorbic acid, which is the reducing agent of proline hydroxylation in collagen biosynthesis. We performed a series of *in vitro* analyses on 10 mg of marine and BDH sands, and on sponge spicules, tested in 5 ml of a 0.5 M solution of ascorbic acid adjusted to pH 7.8 by 0.1 M NaOH at 20 °C. The amount of each material tested that had been dissolved by this organic acid within 24 h was $2.34 \pm 0.48\%$, $3.11 \pm 0.5\%$ and $2.53 \pm 0.36\%$, respectively. SEM observation of etched surfaces both on sands and spicules confirmed that dissolution had occurred.

We then investigated the release of ascorbic acid from a group of specimens, using its ability to reduce oxidized cytochrome *c*. The specimens were mechanically disturbed with a plastic rod, creating an effect similar to the disturbance resulting from sand settling on the sponge surface. Under these conditions, *C. reniformis* secreted a substance capable of reducing cytochrome *c* (Fig. 4). This reduction was partly prevented by the addition in some specimens of ascorbate oxidase, an enzyme able to convert ascorbic acid to its oxidized form, the dehydroascorbic acid (Fig. 4). No reduction was observed in undisturbed animals. Conversely, superoxide dismutase (SOD), an enzyme commonly used to inhibit cytochrome *c* reduction by O_2^- had no effect. These data support the hypothesis that ascorbic acid is responsible for cytochrome *c* reduction, and consequently for the rapid dissolution of quartz into *Chondrosia* ectosome. However, it is

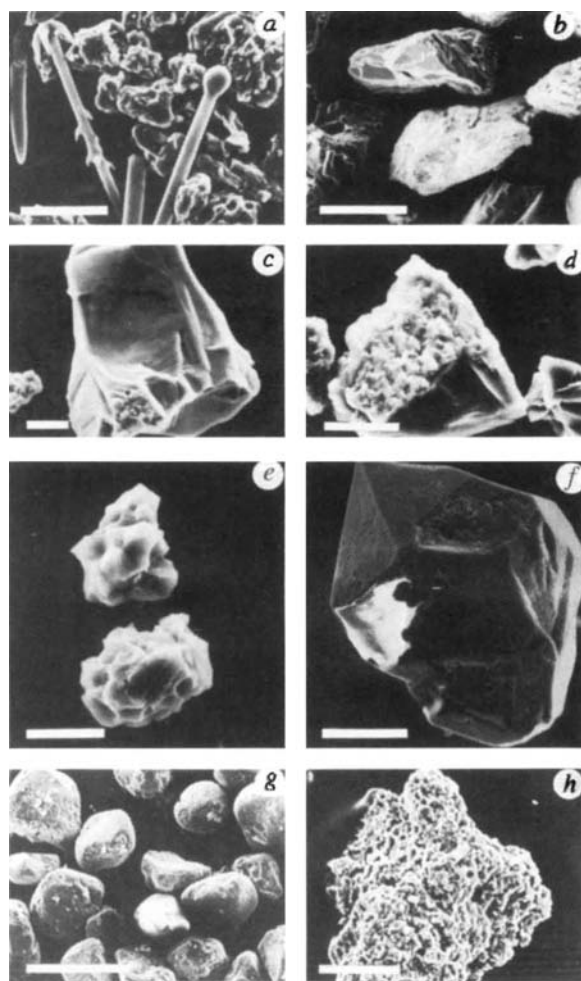


FIG. 1 SEM micrographs showing the action of *C. reniformis* on quartz crystals. a, Siliceous material naturally present in the ectosome of *Chondrosia reniformis*. The quartz is etched, whereas the opaline sponge spicule fragments are unaltered. Scale bar, 50 μm . b, Detail of the marine quartz sand used in the uptake experiment. Scale bar, 100 μm . c–e, Temporal sequence showing progressive quartz dissolution, 1, 2 and 8 days after uptake. Scale bar, 20 μm . f, A partly incorporated monocrystalline bipyramidal quartz. The upper rhombohedra {1011} and {0111}, still external to the sponge ectosome, are intact, whereas a few hours after incorporation the prism {1010} and lower rhombohedra {1011} and {0111} show corrosion features. Scale bar, 500 μm . g, Detail of BDH quartz sand used in the uptake experiment. Scale bar, 500 μm . h, A grain of the same sand eight days after uptake. Scale bar, 10 μm .

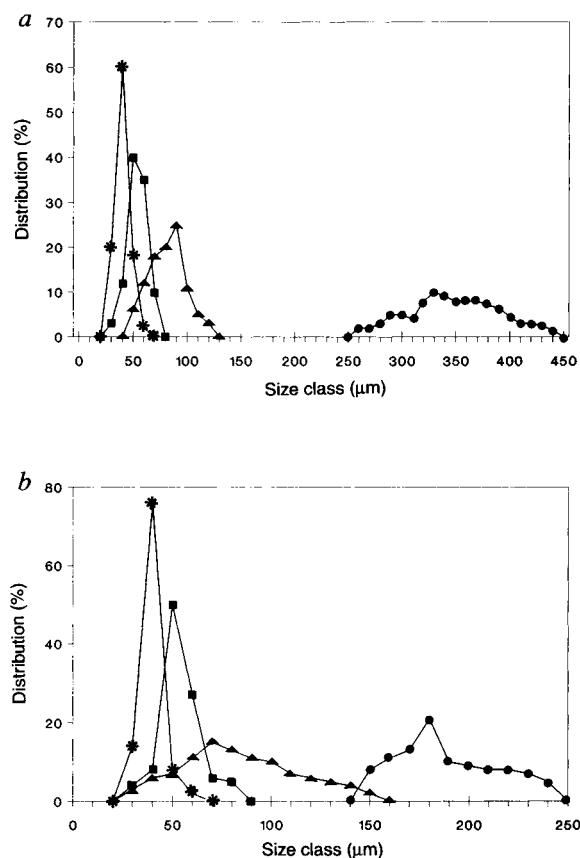


FIG. 2 Size frequency distributions of the maximal diameter of the marine (a) and BDH quartz sand (b) before (circles) and 1 (triangles), 4 (squares) and 8 days (asterisks) after the incorporation in the ectosome of *C. reniformis*.

METHODS. Specimens were collected along the cliffs of the Portofino Promontory and Laigueglia (Ligurian Sea, Italy), and reared at a temperature of 20 °C in 200 l of filtered natural sea water, with an average salinity of 37‰ and pH 7.5.

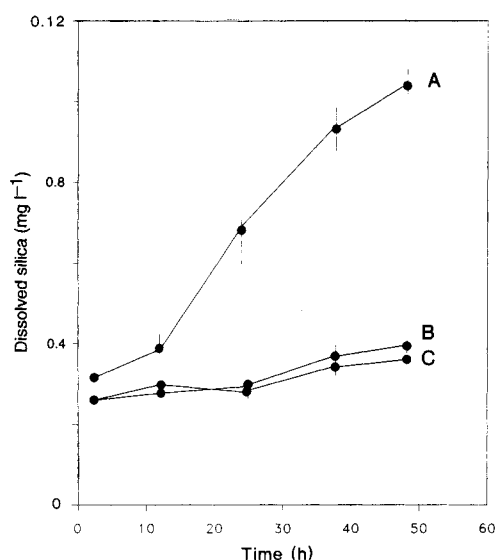


FIG. 3 Dissolution activity of various specimens of *C. reniformis* on BDH sand laid down on their ectosome. a, Trend of the average ($\pm$ s.e.) of dissolution tests on five specimens of *Chondrosia reniformis*, each having been exposed to 1 g of quartz sand settlement. Trends of dissolved silica in three replicates in both ASW with 1 g of dispersed sand (B) and of untreated sponges (C).

METHODS. Specimens with an average wet mass of about 20 g were separately reared at 20 °C in 1 l ASW (420 mM NaCl, 9.3 mM NaCl, 9.3 mM KCl, 25.6 mM MgSO_4 , 22.6 mM MgCl_2 , 9 mM NaHCO_3), aerated by bubbling in plastic containers. ASW was replaced regularly until the content of soluble silica deriving from the residual sponge activity was undetectable. Presence of dissolved silica was checked after 0, 13, 24, 37 and 48 h, by the molybdosilicate method¹³.

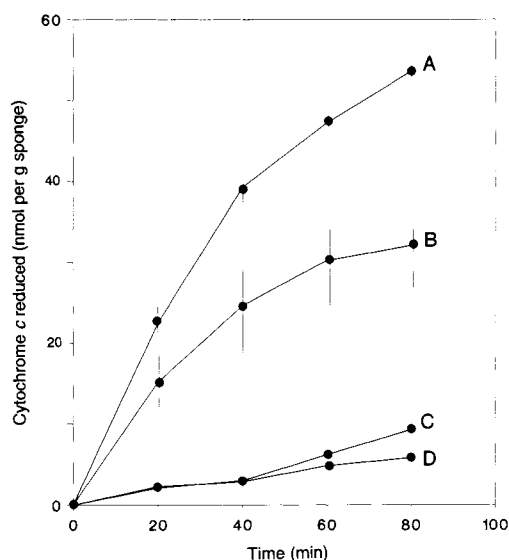


FIG. 4 Production of reducing equivalents by *C. reniformis* with time. Points represent the average $\pm$ s.e. of three experiments. A, Disturbed sponges; B, disturbed sponges in the presence of 20 U ml^{-1} ascorbate oxydase; C, undisturbed sponges; D, undisturbed sponges in the presence of 20 U ml^{-1} ascorbate oxydase.

METHODS. Specimens of mass of about 2 g were placed in 25 ml silica-free ASW (20 °C, pH 7.8), containing 15 μM oxidized cytochrome c.

possible that other reducing agents released from the sponge are also involved.

Hydrated forms of silica, such as opaline spicules, are easy etched *in vitro* by the acid, but they remain completely unaltered within the sponge body. This remarkable selectivity, which cannot be explained by mere chemical processes, is probably due to the cellular mechanisms of particle recognition and compartmentation.

Our data indicate that such action does not induce a complete dissolution of the whole crystal, but seems to work along a definite reticulum, crumbling the grain. The amount of dissolved silica found in the medium is too low to justify a complete dissolution of grains, which otherwise tend to be a uniform size, 25–50 μm (Fig. 2). Carbonate-boring sponges fragment the substrate to chips of the same size and finally eliminate them through their canal system¹.

A complex quartz-dissolving process, such as that described here, is likely to have some physiological value. Because sponges can select the incorporated materials, selective uptake of quartz particles seems to be advantageous to *C. reniformis*. Such grains strengthen the sponge structure, but their packing in the choanosome might clog its vascular apparatus. Particle dissolution could thus be a mechanism to support a continuous input and turnover of readily available quartz crystals. Quartz dissolution by benthic organisms could be a mechanism involved in silica turnover in the marine environment, making dissolved silica available from a previously crystalline condition. □

Received 24 July; accepted 8 September 1995.

1. Rützler, K. & Rieger, G. *Mar. Biol.* **21**, 144–162 (1973).
2. Pomponi, S. A. *Int. Rev. Cytol.* **65**, 301–319 (1980).
3. Pomponi, S. A. *J. mar. biol. Ass. U.K.* **59**, 777–784 (1979).
4. Ascaso, C. & Galvan, J. *Pedobiologia* **16**, 321–331 (1976).
5. Hailbauer, D. K. & Jahns, H. M. *Lichenologist* **9**, 119–122 (1977).
6. Schwab, D. W. & Shore, R. E. *Biol. Bull.* **140**, 125–136 (1971).
7. Sollas, I. B. *J. Ann. Mag. nat. Hist.* **1**, 395–401 (1908).
8. Shaw, E. *Ann. Mag. nat. Hist.* **19**, 601–609 (1927).
9. Schulze, F. E. *Z. wiss. Zool.* **24**, 87–122 (1877).
10. Wiedenmayer, F. *Experientia* **28** (Suppl.) 1–287 (1977).
11. Garrone, R. *Phylogeneses of Connective Tissue* (Karger, Basel and New York, 1978).
12. Bennett, P. C. *Geochim. cosmochim. Acta* **55**, 1781–1797 (1991).
13. American Public Health Association. *Standard Methods for the Examination of Water and Wastewater* 17th edn (APHA, Washington DC, 1989).

ACKNOWLEDGEMENTS. We thank A. De Flora and F. Boero for comments. This work was partly supported by Italian University funds.

A recent common ancestry for human Y chromosomes

Michael F. Hammer

Laboratory of Molecular Systematics and Evolution, Biosciences West, University of Arizona, Tucson, Arizona 85721, USA

THE male-specific portion of the Y chromosome is especially useful for studies of human origins. Patterns of nucleotide variation that are neutral with respect to fitness should permit estimates of when and where ancestral Y chromosomes existed¹. However, variation on the human Y chromosome has been observed to be greatly reduced relative to the autosomes and the X chromosome^{2–5}. One explanation is that selection for a favourable mutation on the non-recombining portion of the Y chromosome has resulted in the recent fixation of a single Y haplotype^{5,6}. A 2.6-kilobase fragment encompassing a polymorphic Alu insertion was sequenced from 16 human and four chimpanzee Y chromosomes. Patterns of nucleotide sequence diversity and divergence provide no evidence for a recent, strong selective sweep on the human Y chromosome. The time back to a common ancestral human Y chromosome is estimated to be 188,000 years, with a 95% confidence interval from 51,000 to 411,000 years. These results are consistent with autosomal and mitochondrial DNA studies that suggest a long-term

human effective population size of 10,000 and a sex ratio of 1 (ref. 7). These inferences contradict predictions of the multiregional hypothesis⁸ positing a widespread transformation of *Homo erectus* populations into *Homo sapiens*.

The results of a sequencing survey of eight Africans, two Australians, three Japanese, two Europeans and four chimpanzees (Table 1) are summarized in Fig. 1. These data strongly suggest that the Y Alu polymorphic (YAP) element inserted on the human Y chromosome a single time after the divergence of humans and chimpanzees. There is an average of one-nucleotide-site difference between two randomly chosen 2.6-kb sequences (YAP locus), or a nucleotide diversity (π) of 0.037% (Table 2).

The approach of Hudson *et al.*⁹ was used to test for selection at the YAP locus. This method (HKA test) is based on the neutral expectation of a correspondence between levels of polymorphism within a species and divergence between closely related species. By comparing these parameters at two unlinked loci, the HKA test identifies selective differences between regions and allows one to distinguish a reduction in polymorphism due to a selective sweep from one due to a population bottleneck. The ratio of nucleotide diversity within humans (π) to divergence between human and chimpanzee (D) at the YAP locus is 1:51. This is comparable to $\pi:D$ ratios at class I sites within the mitochondrial (mt)DNA COII (1:54) and ND4-5 (1:40) loci (Table 2). When the HKA test is applied, the null hypothesis of neutral molecular evolution is not rejected (YAP and COII: HKA test statistic = 0.63, $P=0.43$; YAP and ND4-5: HKA test statistic = 1.03, $P=0.31$). The power of the test was examined by assuming that there were no segregating sites at the YAP locus. The null hypothesis was rejected below the 5% level in both sets of comparisons.

An alternative to a neutral molecular evolution hypothesis is coincidental selective sweeps at both Y chromosomal and mitochondrial loci. There are two lines of evidence against the latter hypothesis. First, a comparison of $\pi:D$ ratios at the YAP and β -globin (1:11) loci (Table 2) is not significantly different from the fourfold difference expected under a neutral model, assuming constant population size and a sex ratio of 1 (HKA test statistic = 0.04, $P=0.85$). Second, average estimates of π and D at silent sites from a large sampling of genes throughout the genome also yield a ratio (1:13) that is fourfold higher than the $\pi:D$ ratio at the YAP locus (Table 2). Therefore, the data provide no evidence for a recent fixation of a favourable mutation anywhere on the non-recombining portion of the Y chromosome (or the mtDNA molecule).

These results are not inconsistent with a recent sequencing survey of the third intron of the *Zfy* gene that revealed complete

TABLE 1 Geographic origin of the individuals and their YAP haplotypes

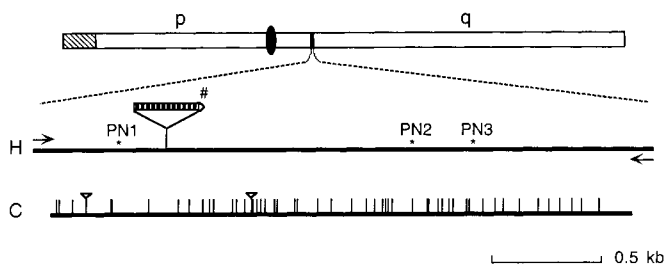
Geographic origin	YAP	3' tail	PN1	PN2	PN3	Haplotype no.
<i>H. sapiens sapiens</i>						
South Africa Xhosa	+	L	C	C	G	3
South Africa Zulu	+	S	T	T	G	5
South Africa Sotho	—	—	C	C	G	1
South Africa !Kung	—	—	C	C	A	2
Zaire Mbuti Pygmy	+	S	T	T	G	5
Zaire Mbuti Pygmy	—	—	C	C	G	1
African American	+	S	T	T	G	5
African American	+	S	T	T	G	5
Australian Aboriginal	—	—	C	C	G	1
Australian Aboriginal	+	S	C	T	G	4
European	—	—	C	C	G	1
European	—	—	C	C	G	1
Japanese	+	L	C	C	G	3
Japanese	—	—	C	C	G	1
Japanese	+	L	C	C	G	3
Cell line 811	—	—	C	C	G	1
<i>P. troglodytes</i>						
Africa (WHT1284)	—	—	C	C	G	—
Africa (Shadow)	—	—	C	C	G	—
Africa (Willy)	—	—	C	C	G	—
Africa (Pacer)	—	—	C	C	G	—

YAP refers to the presence or absence of the YAP element; 3' tail refers to the length of the poly(A) tail (L, long; S, short); and PN1–PN3 refer to the nucleotide states at positions 338, 1,682 and 1,926 in Fig. 1. Haplotype 4 was also identified from partial sequences of YAP⁺ DNA isolated from European samples (data not shown). The following human DNA samples were provided: Xhosa, Zulu and Sotho by T. Jenkins; Pygmies and Japanese by K. Kidd and J. Kidd; an African American and Aboriginal Australians by A. C. Wilson. The !Kung DNA was purchased from the Coriell Institute (Camden NJ) and cell line 811 from American Type Culture Collection (Rockville, MD). The remaining human DNAs were isolated from blood donated by volunteers. D. Page provided WHT1284 DNA; R. Ferrell provided the remaining chimpanzee DNAs.

monomorphism⁵. When the HKA test is applied (in comparisons with both mtDNA and β -globin), the null hypothesis is not rejected. However, the unusually low level of divergence between human and chimpanzee *Zfy* introns (a nearly threefold lower D value compared with the YAP locus) makes this region uninformative for inferring human population history.

The polymorphic nucleotide sites and YAP element define five Y-chromosome haplotypes (Table 1 and Fig. 2). Following the model of Tajima¹⁰, the pair of sequences representing the maximum pairwise divergence for a sample of sequences estimates the expected total divergence. The age of the ancestral human Y chromosome is estimated to be 188,000 years (Table 2 and basal node of Fig. 2), with a standard deviation of 94,000 years (equation (2) of ref. 11). The coalescence time has 95% confidence limits of 51,000 years to 411,000 years (gamma

FIG. 1 Schematic representation of the human Y chromosome and the 2.6-kb YAP region (locus) located on the long arm of the Y chromosome (Yq11). The pseudoautosomal region is indicated by a hatched box on the tip of the short arm (Yp), and the centromere is indicated by a shaded oval. Three polymorphic nucleotide (PN) sites identified in 16 human (H) sequences are indicated by asterisks at positions 338 (PN1), 1,682 (PN2) and 1,926 (PN3). The 300-bp YAP element was found to be identical in sequence and inserted between the same two base pairs (bases 512 and 513) in all eight human YAP⁺ chromosomes examined (see Table 1). The poly(A) tail on the 3' end of the Alu sequence was variable in length (#) with a run of 28 (S) or 45 (L) adenine residues. The YAP regions from four chimpanzee (C) sequences were found to be identical in sequence. There were 50 fixed nucleotide differences between human and chimpanzee sequences: 26 transitions (13 A-G and 13 C-T) and 24 transversions (3 A-C, 7 G-T, 9 A-T and 5 G-C). None of the chimpanzees in the sequence survey, or 14 additional ones tested by PCR, was found to contain the YAP element. In addition to the point mutational and the YAP element insertional differences between humans and chimpanzees, two single-base insertion/deletions (indel; triangles) were identified. Arrows, PCR oligonucleotide.



METHODS. Genomic DNAs were amplified with the primers R1 (5'-AGG-AAGAATACTAGTTTCCT-3') and F1 (5'-GGGGTTTATACTGACCTGCC-3') in a 100- μ l final volume containing 125 ng genomic DNA, 0.12 μ M each primer, 0.2 mM each dNTP, 2.5 mM $MgCl_2$, 50 mM KCl, 10 mM Tris-HCl pH 8.3 and 0.5 U Ampli Taq DNA polymerase (Perkin-Elmer). The cycling conditions were 94 °C for 2 min, and then 30 cycles of 94 °C for 1 min, 55 °C for 1 min and 72 °C for 3 min. Sequencing of the PCR products was carried out as in ref. 14.

TABLE 2 Diversity and divergence at Y chromosomal, mitochondrial and autosomal loci

Locus (n)	Length* (L)	π † (%)	θ ‡ (%)	D§ (%)	$\pi:D$	$T $ ($\times 10^3$)	N_e ¶	Ref.
Y chromosome								
YAP (16)	2,638	0.037	0.034	1.9	1:51	188	4,900	This study
MtDNA								
COII (6)	228	0.73	0.96	39.9	1:54	189	4,600	21
ND4-5 (7)	232	0.90	1.1	35.8	1:40	214	6,300	22
Autosomes								
β -Globin (48)	3,000	0.143	0.13	1.6	1:11	1,100	11,200	18
Average across loci #	—	0.11	—	1.4	1:13	—	10,600	23, this study

* L is the number of noncoding or synonymous sites, or third positions in the case of mtDNA.

† Nucleotide diversity calculated using equation (10.6) of Nei²⁴.

‡ Expected heterozygosity calculated from the number of segregating sites using equation (1) of Kreitman and Hudson²⁵.

§ Average per cent divergence between human and chimpanzee sequences. The values are corrected for multiple substitutions²¹.

|| The coalescence time (T) in years was calculated using equation (1) of Templeton¹¹: $T = \theta(1 + K_{max})/[2\mu(1 + L\theta)]$, where K_{max} is the pairwise difference between the two most divergent sequences in the human clade (Fig. 2), and μ is the mutation rate assuming a human-chimpanzee divergence time of 5 million years¹³: μ for the YAP region is 1.9×10^{-9} sites per year; K_{max} equals 3, 4, 4 and 13 (R. Harding, personal communication) for the YAP region, COII, ND4-5 and β -globin, respectively.

¶ For Y chromosome and mtDNA data, $\pi = 2N\mu_g$, where N is the number of breeding males or females, respectively, and g is the generation time ($g = 15-20$ yr). For autosomal data, $\pi = 4N_e\mu_g$, where N_e is the total effective population size⁷.

Average π at fourfold-degenerate sites estimated from 49 loci in humans²³. Average D at fourfold-degenerate sites estimated from 21 human-chimpanzee gene pairs. The GenBank (release 88.0) databank was screened for all chimpanzee gene sequences and then a similarity analysis program employing the method of Pearson and Lipman²⁶ was used to identify and align human and chimpanzee sequences. A total of 26 fourfold-degenerate site differences were found at 1,826 possible sites for the 21 loci (data available on request).

distribution¹¹). This age agrees with estimates of the ancestral human mtDNA molecule^{12,13}. An implication of these results (Fig. 2) is that the current distribution of Y chromosomes carrying the YAP element^{14,15} reflects migrations of male *H. sapiens* from Africa to Europe and Japan.

A coherent picture now emerges that places both our common ancestral Y chromosome and mitochondrial DNA molecule in the late middle Pleistocene only slightly before the hypothesized

origin of anatomically modern humans based on fossil data¹⁶. The relatively recent ancestry of these genomic counterparts is unlikely to reflect the recent spread of favourable mutations¹⁷. The estimate of a long-term male effective population size of $\sim 5,000$ (Table 2) is in agreement with previous estimates based on nuclear and mtDNA loci^{7,18}. This estimate contradicts predictions of the multiregional hypothesis⁸ because vast regions of Europe, Africa and Asia could not have been continuously inhabited by so few individuals¹⁹. The overall concordance of information from different genomic components (Table 2) increases confidence in recent replacement models^{16,20}. Continued sampling of Y chromosome sequence variation will further clarify the degree and kind of selection operating on the Y chromosome, the extent of geographic subdivision of early human populations, the human male effective population size in the late middle Pleistocene, and the geographic homeland of ancestral Y chromosomes. □

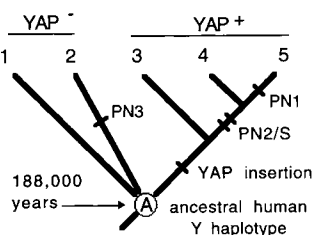


FIG. 2 Evolutionary tree of five human YAP haplotypes. Mutational changes are indicated by a bar across the branch of the tree. The ancestral state at each polymorphic site was determined by comparison with the chimpanzee YAP sequence. Haplotype 1 has undergone no change from the ancestral human Y haplotype (A); haplotype 2 has undergone a single G-to-A transition at PN3. The remainder of the mutations have occurred on the lineage leading to YAP⁺ haplotypes. The first event on this lineage was the insertion of the YAP element with a long poly(A) tail. Subsequently, there was a C-to-T transition at PN2 and a deletion in the poly(A) tail to the S allele (S). Finally, a second C-to-T transition occurred at PN1 giving rise to haplotype 5. The right-hand branch (YAP⁺ haplotypes 3-5) has a coalescence time of 141,000 years with a standard deviation of 81,000 years. The coalescence time has 95% confidence limits of 29,000 years to 34,000 years (see text). This estimate is incompatible with the hypothesis that the YAP element inserted on the Y chromosome of a male *H. erectus* ≥ 1 million years ago.

METHODS. A parsimony analysis using the chimpanzee YAP sequence as an outgroup yielded a single tree with six steps and a consistency index of 1.0. Estimates of the maximum number of sequence differences (K_{max}) among haplotypes on each side of an ancestral node included only nucleotide site differences. K_{max} for all haplotypes was determined by assuming that haplotype 2 attached to the lineage leading to haplotype 1 with the ancestral node separating haplotypes 1 and 2 from haplotypes 3-5. For example, K_{max} for all YAP⁺ haplotypes is 2 (sequence difference between haplotypes 3 and 5) and K_{max} for the whole tree is 3 (sequence difference between haplotypes 2 and 5). This yielded a maximum K_{max} and an upper bound for the estimated age of the ancestral human Y haplotype.

Received 18 May; accepted 15 September 1995.

- Maynard-Smith, J. *Nature* **344**, 591-592 (1990).
- Jakubiczka, S., Arriemann, J., Cooke, H. J., Krawczak, M. & Schmidtke, J. *Hum. Genet.* **84**, 86-88 (1989).
- Ellis, N., Yen, P., Neiswanger, K., Shapiro, L. J. & Goodfellow, P. N. *Cell* **63**, 977-986 (1990).
- Malaspina, P. et al. *Ann. hum. Genet.* **54**, 297-305 (1990).
- Dorit, R. L., Akashi, H. & Gilbert, W. *Science* **268**, 1183-1185 (1995).
- Whitfield, L. S. R., Lovell Badge, R. & Goodfellow, P. N. *Nature* **364**, 713-715 (1993).
- Takahata, N. *Molec. Biol. Evol.* **10**, 2-22 (1993).
- Wolpoff, M. H. in *The Human Revolution: Behavioural and Biological Perspectives on the Origins of Modern Humans* (eds Mellars, P. & Stringer, C.) 123-154 (Edinburgh Univ. Press, Edinburgh, 1989).
- Hudson, R. R., Kreitman, M. & Aguade, M. *Genetics* **116**, 153-159 (1987).
- Tajima, F. *Genetics* **105**, 437-460 (1983).
- Templeton, A. R. *Am. Anthropol.* **95**, 51-72 (1993).
- Stoneking, M. *Evol. Anthropol.* **2**, 60-73 (1993).
- Horai, S., Hayasaka, K., Kondo, R., Tsugane, K. & Takahata, N. *Proc. natn. Acad. Sci. U.S.A.* **92**, 532-536 (1995).
- Hammer, M. F. *Molec. Biol. Evol.* **11**, 749-761 (1994).
- Hammer, M. F. & Horai, S. *Am. J. hum. Genet.* **56**, 951-962 (1995).
- Stringer, C. B. & Andrews, P. *Science* **258**, 1263-1268 (1988).
- Goldman, N. & Barton, N. H. *Nature* **367**, 440-441 (1992).
- Fullerton, S. M., Harding, R. M., Boyce, A. J. & Clegg, J. B. *Proc. natn. Acad. Sci. U.S.A.* **91**, 1805-1809 (1994).
- Rogers, A. & Harpending, H. C. *Molec. Biol. Evol.* **9**, 552-569 (1992).
- Harpending, H. C., Sherry, S. T. & Rogers, A. R. *Curr. Anthropol.* **34**, 483-496 (1993).
- Ruvolo, M. et al. *Molec. Biol. Evol.* **10**, 1115-1135 (1993).
- Kocher, T. D. & Wilson, A. C. in *Evolution of Life: Fossils, Molecules and Culture* (eds Osawa, S. & Honjo, T.) 391-413 (Springer, Tokyo, 1991).
- Li, W.-H. & Sadler, L. A. *Genetics* **129**, 513-523 (1991).
- Nei, M. *Molecular Evolutionary Genetics* (Columbia Univ. Press, New York, 1987).
- Kreitman, M. & Hudson, R. *Genetics* **127**, 565-582 (1991).
- Pearson, W. R. & Lipman, D. J. *Proc. natn. Acad. Sci. U.S.A.* **85**, 2444-2448 (1988).

ACKNOWLEDGEMENTS. I thank S. Zegura, M. Nachman, R. Harding, R. Hudson, M. Wojciechowski, J. Silva, D. Rowe and S. Kaplan for discussions and for comments on the manuscript.

Sequence variation of the human Y chromosome

L. Simon Whitfield*, John E. Sulston†
& Peter N. Goodfellow*

* Department of Genetics, University of Cambridge,
Cambridge CB2 3EH, UK

† Sanger Centre, Hinxton Hall, Hinxton, Cambridgeshire CB10 1RQ,
UK

WE have generated over 100 kilobases of sequence from the non-recombining portion of the Y chromosomes from five humans and one common chimpanzee. The human subjects were chosen to match the earliest branches of the human mitochondrial tree. The survey of 18.3 kilobases from each human detected only three sites at which substitutions were present, whereas the human and chimpanzee sequences showed 1.3% divergence. The coalescence time estimated from our Y chromosome sample is more recent than that of the mitochondrial genome. A recent coalescence time for the Y chromosome could have been caused by the selected sweep of an advantageous Y chromosome or extensive migration of human males.

It has been suspected for some time that the non-recombining portion of the human Y chromosome might harbour a low level of variation compared with the autosomes and the X chromosome. However, studies of restriction-enzyme site polymorphism have not included determinations of the interspecies divergence needed to correct for the mutation rate of the region examined^{1,2}. The amount of divergence between two sequences depends on the time since their last common ancestor and the rate of mutation. More recent comparisons of Y chromosome and autosomal or X chromosome DNA sequence have incorporated an estimate of mutation rate but have not allowed for the different modes of inheritance and levels of recombination^{3,4}.

The Y chromosome region that we have surveyed lies immediately proximal to the Yp pseudoautosomal boundary. It ends at the boundary and includes *SRY*, the mammalian testis-determining gene. This sequence has been extensively investigated to confirm that it contains no coding sequences other than *SRY*⁵. The sequences at the three sites that show substitutions in humans are presented in Table 1 and the genealogy that this suggests is depicted in Fig. 1.

Interspecific comparison was made using 15,680 base pairs (bp) of sequence (after the removal of insertions and deletions) from the same region of the chimpanzee Y chromosome. 207 (1.32%) of these 15,680 bp differ by substitution between the chimpanzee and human sequence cAMF. The *SRY* coding sequence was removed before the calculation of the number of substitutions per site. This does not eliminate any of the three sites found to be variable in humans.

The non-recombining region of the Y chromosome has properties similar to those of the mitochondrial genome (mtDNA). Neither the mtDNA nor the Y-specific portion of the Y chromosome undergoes conventional meiotic recombination and both the Y chromosome and the mitochondrion are inherited from one sex only: the mtDNA is inherited maternally and the Y chromosome patrilineally. We have compared the variation of the human Y chromosome population with that of the mtDNA in a manner that corrects for the different respective mutation rates; a coalescence time for the Y chromosome sample has been calculated using the method of Tamura and Nei⁶, and this has been compared to an estimate of a mtDNA coalescence time made using the same methodology (Table 2)⁶. The method of Tamura and Nei for estimating the numbers of substitutions between pairs of sequences allows for variation of the substitution rate over sites: ignoring such variation, where it exists, can drastically affect estimates of distance. Assuming that the Homo

TABLE 1 Nucleotides in human and chimpanzee subjects at three sites on the Y chromosome

Subjects	Origin	Language	Nucleotide position		
			4,064	9,138	10,831
cAMF	European	Italian	G	C	A
HG2260	Melanesian	Nasioi	G	T	G
HG2264	Rondonian surui	Tupi	G	C	G
HG2486	Tsumkwe san	!Kung	G	C	A
HG2258	Mbuti pygmy	Niger/Kordofanian	A	C	G
Chimpanzee			G	C	A

The names, origins and languages of the human subjects are given, together with the nucleotide that each human and the chimpanzee show at the three positions that are variable in the human sequences. Numbering refers to the sequence of cosmid cAMF3.1 (ref. 5). Sequences with names prefixed HG were obtained from direct sequencing of PCR products; HG numbers are accession numbers of lymphoblastoid cell lines (in the Y Chromosome Consortium repository) from which PCR-template DNA was prepared. cAMF is from the cosmid clone cAMF3.1, the sequencing of which is described in ref. 5. The sequence of cAMF3.1 was used to design PCR primers to produce amplification products covering the whole 18.3-kb region. All PCR primers were tailed at the 5' end with M13 forward or reverse primer sequence. Products from PCR amplifications were purified on agarose gels, then electroeluted, precipitated and washed. Sequencing was by a linear amplification dideoxy termination method using fluorescent dye-labelled primers matching one of the tail sequences⁵. Candidate nucleotide variants were checked by hybridization of radiolabelled allele-specific oligonucleotides to PCR products fixed to nylon membranes (Hybond N+, Amersham). The same PCR primers and methods were used to obtain sequence from the chimpanzee. The lymphoblastoid cell line Colin (S. Marsh and P. Parham) was the source of genomic DNA. The numerous sites at which the chimpanzee sequence differed from the human sequence were not checked with allele-specific oligonucleotides. Some PCR primers did not readily produce template; although a complete sequence contig of the region was produced, only the 15.7 kb of sequence for which both DNA strands were sequenced at least once were used for analysis. This value represents the amount of sequence left after removal of insertions and deletions.

and Pan clades diverged 6 million years before present, the coalescence time of the mtDNA sample is estimated to be between 120,000 and 474,000 years, and the coalescence time of the Y chromosome sample is estimated to be 37,000 and 49,000 years. If the population samples are equivalent, this suggests a significant difference in the population genetics and demography of the Y chromosome and mtDNA.

We have tested three key aspects of the analysis. First, the positioning of the chimpanzee branchpoint during the construction of the human mtDNA tree determines where the human coalescence occurs. Rather than attempt to recover the true phylogeny of the mtDNA dataset, a range of tree reconstruction models were applied and the coalescence times of the trees that they suggested were assessed. The methods employed were maximum parsimony⁷, neighbourhood joining and UPGMA⁸, with several distance estimates, and maximum likelihood⁷ with a variety of transition/transversion ratios. A total of 57 different can-

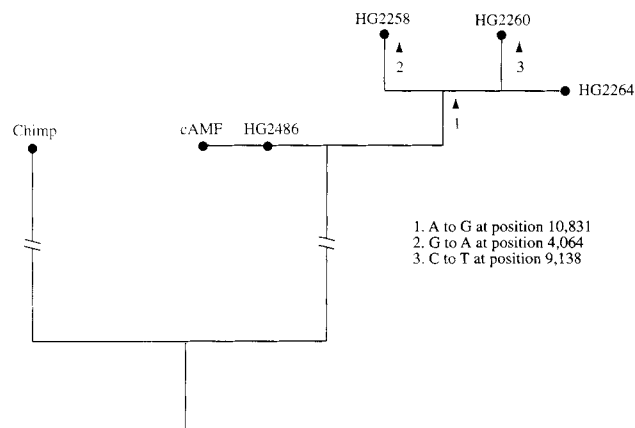


FIG. 1 Genealogy of Y chromosome sequences estimated using the principle of maximum parsimony.

TABLE 2 Estimates of numbers of nucleotide substitutions per site, rate of nucleotide substitution and time taken for human sequences to coalesce

	Y	mtDNA
$\hat{d}_{hh}$ (s.e.)	0.0000942 (0.0000732)	0.024 (0.006)
Mean human-human distance via deepest fork (substitutions per site)		
$\hat{d}_{ch}$ (s.e.)	0.0135 (0.000962)	0.752 (0.224)
Mean chimp-human distance (substitutions per site)		
$\hat{\lambda}_{max}$	1.284×10^{-9}	1.00×10^{-7}
Maximum substitution rate (substitutions per site per year)		
$\hat{\lambda}_{min}$	9.631×10^{-10}	2.53×10^{-8}
Minimum substitution rate (substitutions per site per year)		
Minimum age of coalescence ($\hat{d}_{hh}/2\hat{\lambda}_{max}$)	37,000 yr	120,000 yr
Maximum age of coalescence ($\hat{d}_{hh}/2\hat{\lambda}_{min}$)	49,000 yr	474,000 yr

The approach was as follows: (1) construction of a tree from human and chimpanzee sequences (Fig. 1), which indicates the location of the deepest human fork; (2) estimation of the average number of substitutions per site (using the method described in ref. 6) for all pairwise human comparisons involving individuals from opposite branches of the deepest fork of the human tree ($\hat{d}_{hh}$); (3) estimation of the average number of substitutions per site between chimpanzees and human ($\hat{d}_{ch}$) using the same method, and from this estimate the rate of substitution ($\hat{\lambda}$, substitutions per site per year) as a fraction of the time (T) of the separation of the Homo and Pan clades ($\hat{\lambda} = \hat{d}_{ch}/2T$); (4) use of the estimate of the rate of substitution, $\hat{\lambda}$, to calculate the age of the coalescence time of the human sequences ($\text{Age} = \hat{d}_{hh}/2\hat{\lambda}$). mtDNA values are taken from ref. 6 (but with $\hat{\lambda}$ recalculated using a single value of T). The method models variation of the substitution rate over sites on a γ distribution. Parameter α determines the shape of the γ distribution and hence the degree of variation. Estimation of α is difficult for the Y chromosome data. Yang¹⁶ suggests that $\alpha = 0.2$ may represent extreme variation whereas $\alpha = 0.8$ may represent little variation. Tamura and Nei estimated and used $\alpha = 0.11$ for the mtDNA control region. $\alpha = 0.11$, $\alpha = 2$ and $\alpha = 100$ give very similar coalescence times for our Y data. All values given in the table and in the text are derived with $\alpha = 2$. The ranges of values given for coalescence times arise from the error for the substitution rate $\hat{\lambda}$. The mean of $\hat{d}_{hh}$ is used in this calculation rather than the minimum and maximum; the error for $\hat{d}_{hh}$ is not independent of the error for the substitution rate $\hat{\lambda}$ which has been considered⁶.

didate trees were obtained (after elimination of duplicates). Among these trees were three roots that were different from that used by Tamura and Nei⁶. None made a significant alteration to the date of the mtDNA coalescence.

Second, our Y chromosome sample is small and might be prone to sampling error. It is, however, not a random sample and it is difficult to say, therefore, what is the posterior probability of observing this level of Y chromosome variation. To estimate the number of new variable sites that would have to be observed in a human sample for the estimated Y chromosome coalescence time to overlap that of the mtDNA, we created a hypothetical sequence that has the ancestral sequence at the three known variable sites and then created 'new' substitutions. Whereas we found only three variable sites in five sequences, six new substitutions are required before the resultant coalescence time approached that of the mtDNA (that is, the dataset of the five original human sequences plus the hypothetical sequence gave the Y coalescence time as 90,000–120,000 years).

Third, the method used assumes that the substitution rate is constant in all parts of the tree. This 'molecular clock' assumption was tested for the mtDNA data set (we have insufficient data to test the Y clock) using the program BASEML from the PAML package⁹. A maximum-likelihood fit of the data to the range of reasonable trees was made under a general reversible 'discrete gamma' model¹⁰, with and without the assumption of a molecular clock. The likelihood ratio statistic ($2\Delta l$) from estimates with and without the clock assumption from each of several best trees could then be compared to a χ^2 distribution with 23 degrees of freedom^{11,12}. This analysis indicated that the molecular clock was a reasonable assumption.

The 'globalization' of the human population might be expected to have created isolated subpopulations. True isolation would prevent the homogenization of the whole population by hindering the complete fixation of individual variants. A post-globalization coalescence time for the Y chromosome sample would suggest that there has been recent replacement of Y chromosomes. The conclusion that our sample has a very recent coalescence may be a consequence of our sampling strategy, and the accuracy of the precise coalescence time for the Y chromosome also depends upon the accuracy of our date for the Homo-Pan divergence and upon undetected fluctuations in the molecular clock.

Y chromosome lineages are not shuffled by recombination. When selection acts upon a single Y-linked locus, the whole non-recombining region 'hitchhikes' towards fixation as well^{13–15}. The spread of an individual Y chromosome can occur purely by chance, but it is likely to be greatly influenced by selective advantage, male migration and even reproductive behaviour. For example the practice of polygyny may enable a small number of males to father a disproportionately large number of offspring, and their Y chromosomes could increase in frequency very rapidly. Nonetheless, widespread replacement of Y chromosomes is most readily explained by male migration and the occurrence of a selectively favoured Y chromosome. □

Received 18 September; accepted 14 October 1995.

- Jakubiczka, S., Arnemann, J., Cooke, H. J., Krawczak, M. & Schmidtke, J. *Hum. Genet.* **84**, 86–88 (1989).
- Malaspina, P. et al. *Ann. Hum. Genet.* **54**, 297–305 (1990).
- Dorit, R. L., Akashi, H. & Gilbert, W. *Science* **268**, 1183–1185 (1995).
- Ellis, N. et al. *Nature* **344**, 663–665 (1990).
- Whitfield, L. S., Hawkins, T. L., Goodfellow, P. N. & Sulston, J. *Genomics* **27**, 306–311 (1995).
- Tamura, K. & Nei, M. *Molec. Biol. Evol.* **10**, 512–526 (1993).
- Felsenstein, J. *PHYLP (Phylogeny Inference Package)* (University of Washington, Seattle, USA, 1993).
- Kumar, S., Tamura, K. & Nei, M. *MEGA (Molecular Evolutionary Genetics Analysis)* (Pennsylvania State University, University Park, PA, 1993).
- Yang, Z. *PAML (Phylogenetic Analysis by Maximum Likelihood)* (Inst. Molec. Evol. Genet., Pennsylvania State University, University Park, PA, 1995).
- Yang, Z. B. *J. molec. Evol.* **39**, 105–111 (1994).
- Yang, Z., Goldman, N. & Friday, A. *Syst. Biol.* **44**, 385–400 (1995).
- Felsenstein, J. *J. molec. Evol.* **17**, 368–376 (1981).
- Maynard-Smith, J. & Haigh, J. *Genet. Res.* **23**, 23–35 (1974).
- Begun, D. J. & Aquadro, C. F. *Nature* **356**, 519–520 (1992).
- Clark, A. G. *Genetics* **115**, 569–577 (1977).
- Yang, Z. H. *J. molec. Evol.* **40**, 689–697 (1995).

ACKNOWLEDGEMENTS. We thank N. Barton and J. Barrett for their comments on the manuscript, A. Friday for discussion, N. Ellis and M. Hammer for cells and DNA from the Y Chromosome Consortium repository, and T. Hawkins for his input to the design of the sequencing strategy. This work was supported by grants from the Wellcome Trust (to P.N.G. and J.E.S.). L.S.W. is supported by an Imperial Cancer Research Fund graduate bursary.

Stereoscopic depth perception at high velocities

M. J. Morgan & E. Castet*

Department of Visual Science, Institute of Ophthalmology, Bath Street, London EC1V 9EL, UK

* Laboratoire de Psychophysique, Université Louis Pasteur, Rue Goethe, Strasbourg, France

THE view of the world from different perspectives provided by the two eyes is used by the human visual system to compute the relative distances and solid shapes of objects¹. However, the traditional theory of binocular disparity takes little account of the fact that a moving target will stimulate many different sets of disparate points in the two eyes with a range of temporal delays. Here we show that stereoacuity for periodic gratings is not degraded by velocities of up to 640° s^{-1} provided that they do not move at a greater rate than 30 cycles s^{-1} . The minimum detectable spatial phase difference between the eyes was equivalent to a spatial phase difference of about 5° and an interocular temporal delay as small

as 450 μ s. We suggest that stereopsis for moving targets is accomplished by neurons having a spatial-temporal phase shift in their receptive fields between the eyes.

The traditional theory of binocular stereopsis is based on the concepts of corresponding points and relative disparity. When the two eyes are converged on a single object, the images of that object fall on corresponding points in the two eyes and are defined as having a zero disparity. The images of objects not being fixated fall on the two retinæ with a range of disparities from which their relative distances can in principle be computed. The computation of binocular disparity may be one function of cells in the visual cortex that are driven by inputs from both the left and right eye. Cortical detectors of retinal disparity are thought to have receptive fields in anatomically non-corresponding points of the two eyes²; or, alternatively, receptive fields in the same points but with different spatial phase sensitivities³. A single, stationary target will optimally stimulate a disparity detector when its left and right eye images both fall in the receptive fields of that detector.

The geometry of binocular disparity is usually illustrated by static diagrams. However, in real life both the observer and objects in the scene are likely to be moving, some with high velocities, such as when the hand is rapidly moving to grasp an object. A moving image will stimulate a whole range of disparity-tuned detectors with a corresponding range of temporal delays. Consider, for example, the case of a target moving horizontally with a velocity of 100° s⁻¹ and a disparity of 0.1°. Although this target is never present instantaneously on corresponding retinal points, it will stimulate corresponding points with a delay or phase difference of 1 ms. To distinguish this situation from one in which the target actually has zero disparity, the ensemble of disparity detectors must be able to discriminate the signal arising from a 1 ms phase difference from that arising from a zero temporal phase difference.

The better the temporal resolution of disparity-detecting neurons, the more sensitive they will be to interocular phase differences, and thus to the disparity of a moving target. This might seem to raise a problem, because experiments have shown that the stereo system is temporally sluggish, with a marked fall off if the depth of the stimulus is oscillated at frequencies above 5 Hz (ref. 4). However, these data refer to the response of stereopsis to changes in depth over time, not to detection of a constant disparity at different velocities of the monocular image. To explore the velocity limits of stereopsis we used moving gratings with an interocular phase difference.

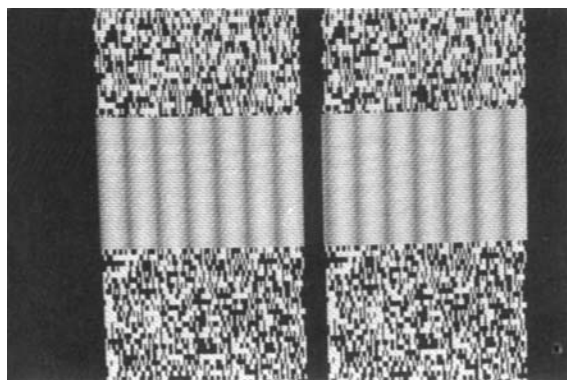
It is known that precise stereoacuity for horizontally moving single bars can be maintained with image velocities of up to 5° s⁻¹ (refs 5, 6), but the upper velocity limits of stereopsis have

not previously been explored. We used horizontally moving vertical sinusoidal gratings as stimuli (Fig. 1), because these permit independent manipulation of velocity (° s⁻¹) and temporal frequency (Hz). The spatial frequency of a grating is the number of cycles of the grating per unit of distance, conventionally expressed as cycles per degree (c.p.d.) of visual angle. The temporal frequency of a moving grating is the rate at which one cycle of the grating moves past a given point in space, expressed in cycles s⁻¹ (Hz). The relation between velocity (v), spatial frequency (sf) and temporal frequency (tf) is described by: $v = tf/sf$. In order to obtain low spatial frequencies and velocities of up to 1,000° s⁻¹ we viewed the screen from a very close distance (14 cm)⁷ and used the lenses from a Brewster stereoscope both to aid accommodation and to fuse the left and right eye images, which were placed side by side on the screen (see Fig. 1).

We found that stereoacuity was actually improved in some cases by movement, as shown by the relative thresholds of less than 1.0 in Fig. 2. This was particularly evident for the lowest spatial frequency used (0.04 c.p.d.). The remarkable fact revealed by the top two panels in Fig. 2 was that stereoacuity could be maintained for very high velocities: as high as 640° s⁻¹ in the case of the lowest spatial frequency. The upper velocity limit was reduced for higher spatial frequencies, and this suggests that the true limit for stereopsis may not be velocity, but rather temporal frequency, because the temporal frequency corresponding to a given velocity of a grating is proportional to its spatial frequency. When thresholds were plotted as a function of temporal frequency (in the bottom panels of Fig. 2) the point of breakdown was indeed similar for all frequencies: in the region of 30–40 Hz. The result is similar to that previously reported for direction discrimination thresholds⁷. However, at our temporal frequency limit for detecting depth, the direction of movement of the grating could still be perceived, so depth and direction thresholds are not identical.

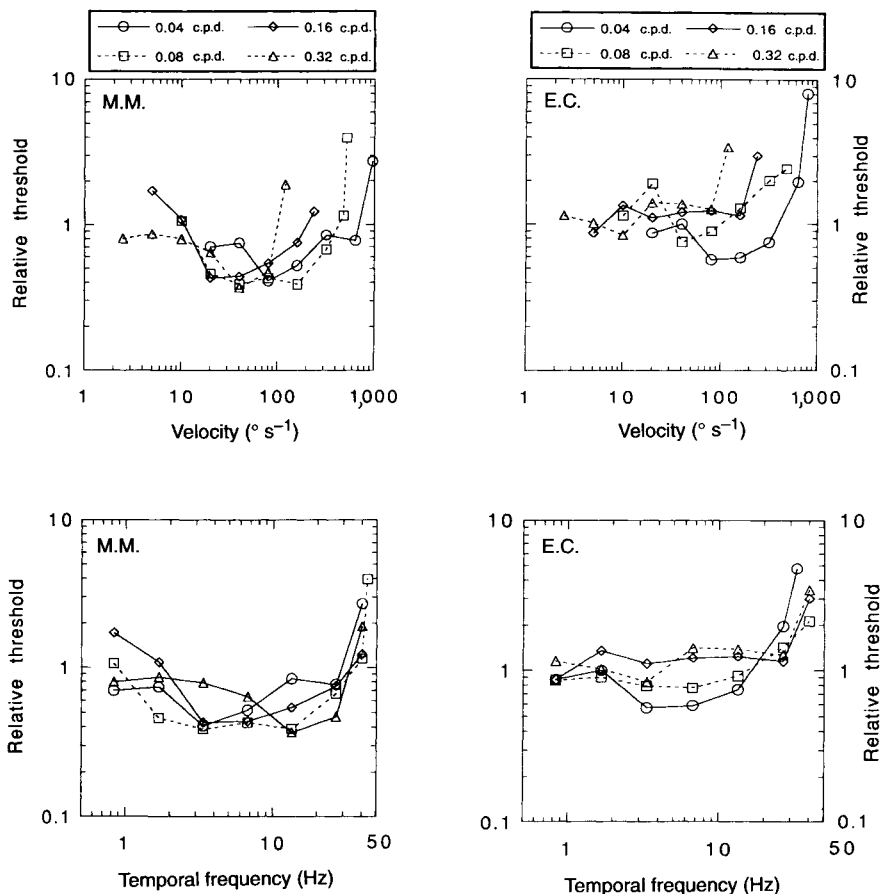
When the stereo thresholds were expressed as inter-ocular delays (that is, as the spatial disparity at threshold divided by the velocity) they showed minimum values of 1 ms or even less (Fig. 3). The lowest threshold found was 450 μ s, similar to that reported for detecting interocular delay in dynamic visual noise⁸. It follows that the neural mechanisms subserving dynamic binocular vision, like those of the auditory system⁹, must be sensitive to temporal phase differences smaller than the inter-spike interval of neurons. Stereo-depth from inter-ocular delay is known as the Pulfrich effect¹⁰ and it has been suggested that the effect may depend directly upon a mechanism for detecting temporal delays, rather than one based on spatial disparities^{11,12}. The present data do not support the idea of a purely temporally based mechanism because the threshold delay was far from con-

FIG. 1 The photograph illustrates stationary versions of the gratings used in the experiments to determine stereoacuity. Observers fused the left and right half images with the help of a Brewster stereoscope positioned 14 cm from the screen, such that 1 screen pixel subtended 0.2° visual angle. The computer-generated patterns were displayed on a Barco Calibrator II monitor with a 200 Hz frame rate. The gratings were moved at the desired velocity by displacing them horizontally the required number of pixels every frame or multiple of frames. The gratings had a mean luminance of 23.5 cd m⁻² and a Michelson contrast ($100 \times (L_{\max} - L_{\min}) / (L_{\max} + L_{\min})$) of 78%. Stereo disparities were introduced by advancing or retarding the phase of the right eye grating relative to the left, which caused the grating to appear either in front of or behind the plane of the surrounding static random dot pattern. The grating was exposed for 0.5 s, preceded and followed by a flash of random noise in the aperture to mask the onset and offset of the grating. The observer's task was to decide on each trial whether the grating was in front of or behind the stationary surround. The threshold was defined as the amount of phase shift needed for the observer to be 84% correct, after subtracting any bias due to making more 'behind' than 'in front' decisions. The 84% correct point corresponds to the standard deviation of the psychometric function, determined by an adaptive psychometric procedure that chose the disparity on each trial



to be around the observer's threshold, as determined by previous performance²³.

FIG. 2 The four panels show the results of measuring stereo thresholds for a moving grating at a range of velocities (top two panels) and at their equivalent temporal frequencies (bottom two panels). The methods for determining stereo thresholds (vertical axis) are described in the legend to Fig. 1. Results are shown separately for the two authors (M.M.: left-hand panels; E.C.: right-hand panels). The four kinds of symbol correspond to four different spatial frequencies of grating, measured as the number of cycles of the grating per degree of visual angle (cycles per degree, c.p.d.). All thresholds are expressed relative to the threshold for a stationary grating: thus values less than unity imply that performance for the moving grating was better than for the stationary. The top two panels show that as velocity increased, acuity either remained the same as for a stationary grating or improved, until a limiting velocity at which performance broke down. At velocities greater than the rightmost point on the graphs, a threshold could not be determined because the observer's responses were random. The limiting velocity increased as the spatial frequency of the grating decreased. However, when performance is expressed as a function of temporal frequency (bottom two panels) it begins to break down at roughly the same point (30–40 Hz) for all spatial frequencies.



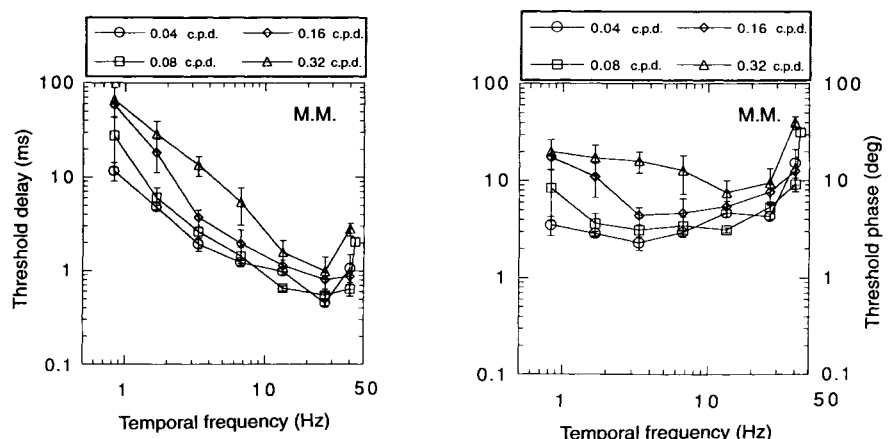
stant over spatial and temporal frequency. In contrast, thresholds expressed as spatial phase differences were more nearly constant, with best values in the region of 5° of phase angle, as previously reported for stationary low spatial frequency gratings¹³.

The simplest explanation of the data would be that interocular spatial phase differences are detected by neural mechanisms with a high temporal frequency response, probably neurons in the magnocellular 'motion' pathway^{14,15}. However, other data are difficult to reconcile with a purely spatial-phase sensitive mechanism. Depth is seen from inter-ocular delay even when all the spatial sampling positions of a discretely moving target are identical in the two eyes¹². Also, inter-ocular differences in the temporal phase of luminance modulation of a single moving bar results in perceived depth even when there is no spatial disparity, an effect that breaks down at temporal frequencies greater than 30–40 Hz (ref. 16), as in the present

experiment. Finally, inter-ocular delay produces depth when viewing purely random dynamic noise⁸, possibly because of the embedding within the noise of horizontal velocity Fourier components¹⁷.

We suggest that stereodisparity tuning for quickly moving targets is carried out by neurons in which the distinction between spatial and temporal disparities is largely obliterated by their spatial-temporal filtering properties. DeAngelis *et al.*³ have recently proposed a model of binocular stereopsis in which binocularly tuned neurons have slightly different spatial-temporal receptive fields in the two eyes. A spatial-temporal receptive field represents the tuning of a cell both in time and in space, and velocity-tuned cells with non-separable profiles tilted in space-time have been inferred from psychophysical^{18,19} and physiological^{20,21} measurements. The way in which spatial and temporal offsets become indistinguishable through a spatial-temporal filter can be demonstrated directly by temporal filtering

FIG. 3 The left-hand panel show the threshold for detecting a disparity between the left- and right-eye gratings transformed into a temporal delay, according to the relation: temporal delay = spatial disparity/velocity. The lowest thresholds are in the region of 450 μ s. Note that the threshold temporal delay is not constant, but decreases with velocity up to the temporal frequency limit of about 30 Hz. The right-hand panel shows the same data expressed as threshold spatial phase. The spatial phase thresholds depend less upon temporal frequency than do the temporal thresholds. Best spatial phase thresholds are in the region of 5° . The vertical bars represent standard deviations of thresholds across repeated measures.



of a motion signal²². DeAngelis *et al.* conceive their cells as having receptive fields with corresponding positions of their envelopes in the two eyes but with a spatial phase difference. Their model can be generalized to include the possibility of temporal phase differences as well, or any combination of spatial

and temporal phase difference, resulting in cells with phase differences in space-time. If such cells are repetitively stimulated at temporal frequencies near to their resolution limit, information about temporal phase will be lost, leading to the effects we have observed in our experiment. □

Received 5 June; accepted 13 September 1995.

1. Wheatstone, C. *Phil. Trans. R. Soc.* **2**, 371–393 (1838).
2. Barlow, H. B., Blakemore, C. & Pettigrew, J. D. *J. Physiol.* **193**, 327–342 (1967).
3. DeAngelis, G. C., Ohzawa, I. & Freeman, R. D. *Nature* **352**, 152–159 (1991).
4. Regan, D. & Beverley, K. I. *Vision Res.* **13**, 2369–2379 (1973).
5. Westheimer, G. & McKee, S. P. *J. opt. Soc. Am.* **68**, 450–455 (1978).
6. Morgan, M. J., Watt, R. J. & McKee, S. P. *Vision Res.* **23**, 541–546 (1983).
7. Anderson, S. J. & Burr, D. C. *Vision Res.* **25**, 1147–1154 (1985).
8. Burr, D. C. & Ross, J. *Vision Res.* **18**, 523–532 (1978).
9. Palmer, A. R., Rees, A. & Caird, D. *Hearing Res.* **50**, 71–86 (1990).
10. Pulfrich, C. *Naturwissenschaften* **10**, 553–564 (1922).
11. Ross, J. *Nature* **248**, 363–364 (1974).
12. Morgan, M. J. & Thompson, P. *Perception* **4**, 3–18 (1975).

13. Schor, C. M., Wood, I. C. & Ogawa, J. *Vision Res.* **24**, 573–578 (1984).
14. Derrington, A. M. & Lennie, P. *J. Physiol. Lond.* **357**, 219–240 (1984).
15. Livingstone, M. S. & Hubel, D. H. *Science* **240**, 740–749 (1988).
16. Morgan, M. J. & Watt, R. J. *Vision Res.* **22**, 863–866 (1982).
17. Morgan, M. J. & Tyler, C. W. *Proc. R. Soc. B* (in the press).
18. Ross, J. & Burr, D. in *Vision, Brain & Cooperative Processes* (eds Arbib, M. & Hanson, A. R.) (Bradford, Amherst, MA, 1983).
19. Burr, D. C., Ross, J. & Morrone, M. C. *Proc. R. Soc. Lond.* **B227**, 249–265 (1986).
20. Reid, R. C., Soodak, R. E. & Shapley, R. M. *J. Neurophysiol.* **66**, 505–529 (1991).
21. McLean, J., Raab, S. & Palmer, L. A. *Visual Neurosci.* **11**, 271–294 (1994).
22. Morgan, M. J. *Perception* **9**, 161–174 (1980).
23. Watt, R. J. & Andrews, D. P. *Curr. Psychol. Rev.* **1**, 205–214 (1981).

ACKNOWLEDGEMENTS. We thank L. Partridge and J. Levitt for helpful comments. Supported by a grant from the BBSRC.

Behavioural abnormalities in male mice lacking neuronal nitric oxide synthase

Randy J. Nelson*, Gregory E. Demas*, Paul L. Huang†, Mark C. Fishman‡, Valina L. Dawson§, Ted M. Dawson¶ & Solomon H. Snyder|¶***

* Department of Psychology, Behavioral Neuroendocrinology Group, The Johns Hopkins University, Baltimore, Maryland 21218-2686, USA

† Cardiovascular Research Center, Massachusetts General Hospital, Boston, Massachusetts 02129, USA

‡ Department of Neurology, § Department of Physiology,

|| Department of Neuroscience, ¶ Departments of Pharmacology and Molecular Sciences and ** Psychiatry and Behavioral Sciences, The Johns Hopkins University, 725 North Wolfe Street, Baltimore, Maryland 21205, USA

In addition to its role in blood vessel^{1,2} and macrophage^{3,4} function, nitric oxide (NO) is a neurotransmitter⁵ found in high densities in emotion-regulating brain regions^{6–8}. Mice with targeted disruption of neuronal NO synthase (nNOS) display grossly normal appearance, locomotor activity, breeding⁹, long-term potentiation¹⁰ and long-term depression¹¹. The nNOS[−] mice are resistant to neural stroke damage following middle cerebral artery ligation¹². Although CO₂-induced cerebral vasodilatation in wild-type mice is NO-dependent, in nNOS[−] mice this vasodilatation is unaffected by NOS inhibitors¹³. Establishing a behavioural role for NO has, until now, not been feasible, as NOS inhibitor drugs can only be administered acutely and because their pronounced effects on blood pressure and other body functions obfuscate behavioural interpretations. We now report a large increase in aggressive behaviour and excess, inappropriate sexual behaviour in nNOS[−] mice.

In establishing our breeding colony of nNOS[−] mice, we housed groups of five nNOS[−] male mice together in cages and upon routine morning examinations often discovered one or two dead mice in each cage. Initial direct observations indicate that the male nNOS[−] mice engage in chronic aggressive behaviour, not apparent among nNOS[−] female mice or wild-type male or female mice housed together (Fig. 1).

To examine this behaviour in greater detail, we conducted studies of inter-male aggression in an intruder-resident model¹⁴. The following aggressive behaviours were scored: offensive attack, biting, wrestling and chasing. Tail rattling was not

recorded as an aggressive behaviour. Submissive postures were also recorded. Submission was operationally defined as rolling onto the back with the paws extended. When wild-type mice are introduced to a cage, the latency to the first aggressive attack by the resident mice is the same for wild-type and nNOS[−] residents (Fig. 2). However, the nNOS[−] residents display 3–4 times more aggressive encounters than wild-type residents. Moreover, in these encounters attacks are initiated by the nNOS[−] residents in 87% of their encounters with the wild-type intruders, nearly six times higher than the proportion of attacks initiated by wild-type residents.

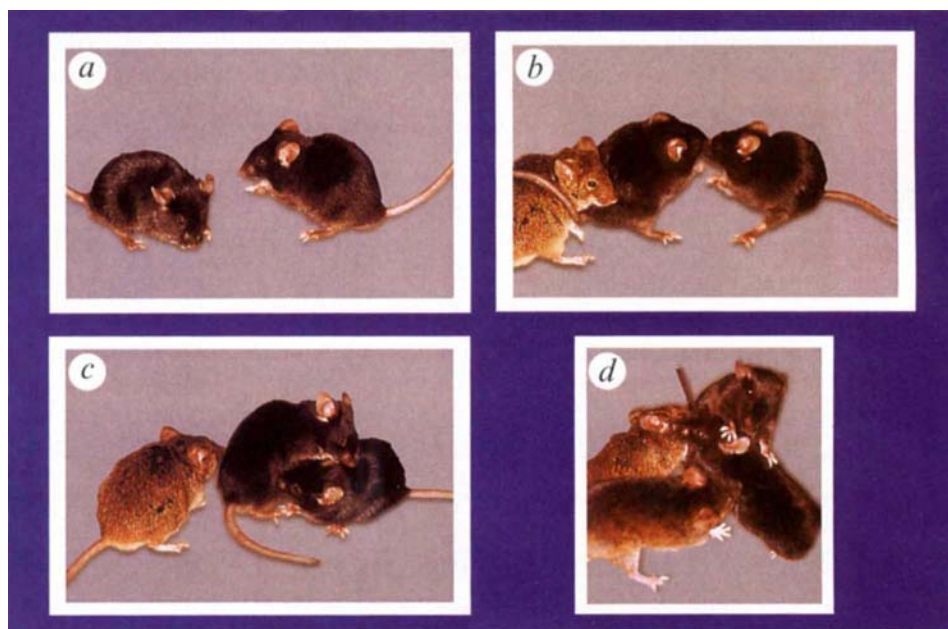
In another model we observed groups of four wild-type or nNOS[−] males together in an aquarium (Fig. 3). In this model the latency to first attack is less than 1 min for the nNOS[−] animals, about 1/5th the latency for wild-type mice. The nNOS[−] mice display almost twice as many attacks as wild-type animals during a 15 min observation period. It is likely that the disparity between the nNOS[−] and wild-type mice would be greater with a longer duration, but we stopped the encounters after 15 min to prevent serious wounding. The duration of aggressive encounters is also greater in nNOS[−] than wild-type mice. The most dramatic difference between the two groups involves the number of submissive postures. nNOS[−] mice display submissive postures only one-tenth as frequently as wild-type mice.

Inappropriate aggressiveness has only been observed in male nNOS[−] mice. In multiple experiments female nNOS[−] mice have not exhibited aggressive behaviour when challenged by an intruder female. Furthermore, female nNOS[−] mice grouped 4 per cage also do not display any aggressive behaviours.

Our initial behavioural observations suggested that the persistence of excessive behaviours seen in aggressive encounters extends to reproductive interactions. In our efforts to obtain timed pregnant females, we paired nNOS[−] males with nNOS[−] females at varying stages of oestrus. The males display excessive and inappropriate mounting behaviour associated with substantial vocal protestations by the females. To quantify this behaviour we conducted studies in which anoestrous females were paired with wild-type or nNOS[−] males commencing at 10:00. Animals were observed for 8 h and their behaviours were recorded by videotape and scored blindly. Two 'live' observers, also uninformed about the phenotype of the mice, scored the behavioural interaction during the first 15 min of each hour (Fig. 4). During the initial 15 min observation period wild-type and nNOS[−] males mount females to the same extent, reflecting the normal male response to a new female. This behaviour diminishes rapidly in wild types so that by 1 h after introduction into the cage the number of mounts by wild-type males decreases 50%. By contrast, only a modest, non-significant decrease in the number of mounts occurs for the nNOS[−] animals. By 7 and 8 h following introduction to the cage, wild-type mouse encounters

*** To whom all correspondence should be addressed.

FIG. 1 Photographs of $n\text{NOS}^-$ male mice engaged in an aggressive encounter. In *a*, the mice are preparing to engage. Both black $n\text{NOS}^-$ mice in *b* are in the classic boxing stance while a brown $n\text{NOS}^-$ mouse looks on. In *c*, the two black $n\text{NOS}^-$ mice are fighting, and neither mouse has displayed a submissive posture. In *d*, the two brown $n\text{NOS}^-$ mice join in the aggressive encounter. These photographs were obtained by simultaneously introducing four $n\text{NOS}^-$ mice into a neutral arena in the experimental model used in Fig. 3. The photographed interactions commenced about 10 min after the mice were placed in the aquarium. The sequence of depicted events occurred over a 2 min period. $n\text{NOS}$ null mice were obtained from a breeding colony established at Johns Hopkins University using animals previously produced by homologous recombination⁹. All animals were screened for homozygosity of the $n\text{NOS}$ null mutation by Southern blot analysis. Wild-type animals consisted of both age-matched C57B6J and SvEv 129 strains to control for possible strain effects because of the $n\text{NOS}$ null mice genetic background. Additional controls consisted of age-matched heterozygous $n\text{NOS}^{+/-}$ littermates who displayed normal behaviour when males were paired together, similar to wild-type mice. Animals were 4–6 months of age at the onset of testing (sexually mature), and were maintained individually in LD 16:8 photoperiods (light on 7:00 EST) at $20 \pm 2^\circ\text{C}$



temperatures and relative humidity of $50 \pm 5\%$ throughout the study. Food (Prolab 1000; Syracuse, NY) and tap water were available ad libitum. All statistical comparisons involved one-way analyses of variance (ANOVA) except where otherwise noted. Pair-wise comparisons were accomplished with Student's *t*-tests. Mean differences were considered statistically significant if $P < 0.05$. Aggression tests were conducted as described previously¹⁴.

have decreased to levels only one-sixth those of the initial period. At these times the number of mounts by $n\text{NOS}^-$ animals continues to be 2–3 times greater than values for wild-type animals.

Both in studies of aggression and of sexual behaviour the $n\text{NOS}^-$ animals display a marked increase in inappropriate aggressive and sexual behaviour reflected in persistent fighting and mounting behaviour despite obvious signals of surrender or

disinterest, respectively, by their test partners. Generally, both aggression and sexual behaviour are enhanced by elevated testosterone levels. Accordingly, we monitored blood testosterone levels in wild-type and $n\text{NOS}^-$ males 2 weeks before any behavioural testing as well as 1 h after completion of the aggression tests. Blood plasma testosterone levels do not differ between the $n\text{NOS}^-$ and the wild-type males at either time point and there is no significant difference in the levels before and after the behavioural study.

Is it possible that sensory or motor abnormalities in the $n\text{NOS}^-$ animals account for the apparent increase in inappropriate aggressive and sexual behaviour? For instance, perhaps the $n\text{NOS}^-$ animals fail to recognize social inhibitory stimuli, such as an odour emitted by a non-oestrous female. Accordingly, we monitored olfactory behaviour by assessing the ability of mice to find a cookie hidden beneath the cage bedding. Latency to discover the hidden cookies does not differ between wild-type and $n\text{NOS}^-$ males and females. Might our behavioural observations be influenced by differences in strength and agility between wild-type and $n\text{NOS}^-$ animals? We monitored the ability of mice to turn around in a blind alley, to walk across a suspended pole, and to hang from a pole (Fig. 4 legend). $n\text{NOS}^-$ males and

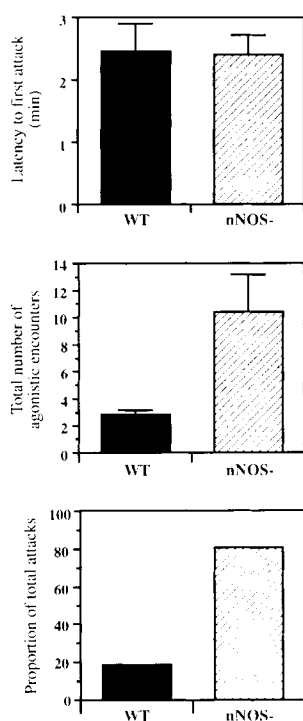


FIG. 2 Defensive aggressive behaviour in $n\text{NOS}^-$ and wild-type mice. These tests were modified from the procedures described previously to assess aggression in genetic knock-out mice^{15,16}. A wild-type intruder was introduced into the home cage of either an $n\text{NOS}^-$ or wild-type adult male mouse ($n = 8$), and the defensive aggressive behaviours were recorded. The latency to first aggressive encounter, the number of agonistic encounters, and the proportion of aggressive behaviours initiated by the resident male during the 5-min tests were recorded. Aggression tests were conducted each day for three consecutive days between 15:00 and 17:00, and a unique pairing was made for each test. The data (mean $\pm$ s.e.m.) were combined from the three tests and plotted in this figure. All behavioural tests were scored blindly.

females do not display any decrease in performance in these measures. Conceivably the nNOS^{-/-} animals are frightened, leading to increased fighting behaviour. We monitored behavioural anxiety in our open field test. We observe no difference between nNOS^{-/-} and wild-type males and females.

Mice with targeted disruption of genes for calcium/calmodulin kinase II (ref. 15) and serotonin receptor 5-HT_{1B}¹⁶ also display enhanced inter-male aggression. Augmented aggression is not likely to be a general feature of animals with targeted disruption of any gene, as animals with deletion of genes for oestrogen receptors exhibit reduced aggressive behaviour¹⁷. In some instances¹⁵ mice with gene deletions display major anatomical and physiological abnormalities and appear 'sick', which can complicate interpretations of discrete behavioural deficits. For instance, mice with deleted monoamine oxidase A display increased intermale aggression but also tremble, have difficulty in righting themselves, are fearful and blind²¹. No neuroanatomical or physiological disturbances have been observed in the nNOS mice whose behaviour superficially appears normal⁹.

FIG. 3 Offensive aggressive behaviour among grouped nNOS^{-/-} and wild-type male mice in a neutral area. Four adult male wild-type or four nNOS^{-/-} ($n=16$) mice were simultaneously introduced into a clear glass aquarium ($38.5 \times 26.5 \times 30.7$ cm). The latency to first aggressive encounter, the duration of each agonistic encounter, the proportion of submissive postures, and the total number of attacks during the 15 min test were recorded for the entire group of four mice. Aggression tests were conducted on two consecutive days between 11:00 and 13:00, and the data (mean $\pm$ s.e.m.) were combined for this figure. All behavioural tests were scored blindly.

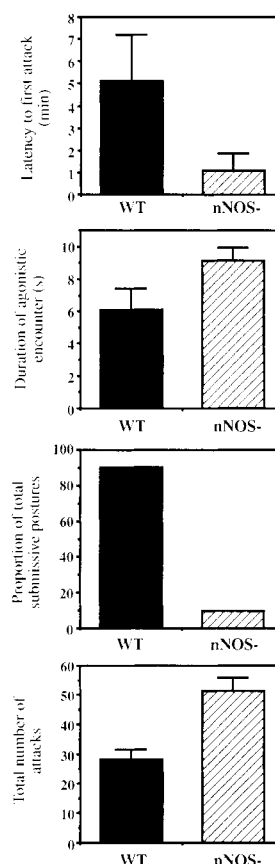
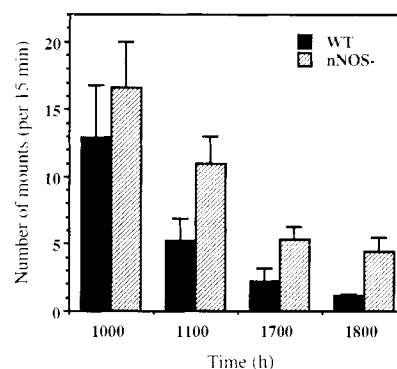


FIG. 4 Sexual behaviour of nNOS^{-/-} and wild-type mice. The mean ($\pm$ s.e.m.) number of mounts (per 15 min observation period) were monitored for nNOS^{-/-} ($n=8$) or wild-type ($n=8$) male mice when paired with a non-oestrous female in a neutral arena. Males were initially placed in a clear glass aquarium ($38.5 \times 26.5 \times 30.7$ cm) for a 15 min acclimatization period. After this a non-oestrous wild-type female was introduced into the mating arena. Animals remained paired for 8 h from 10:00 to 18:00. The number of mounts and aggressive behaviours were recorded by two observers, who were uninformed about the genotype of the mice, during a 15 min observation period at the start of each hour. The behavioural tests were also videotaped continuously for subsequent confirmation of the live scoring. A mount was operationally defined as the male assuming the copulatory position, but failing to achieve intromission. Intromission was defined as the male's penis entering the vagina in association with thrusting behaviour. No animals displayed an intromission. There was only one observed incidence of aggressive behaviour, and it was performed by a nNOS^{-/-} mouse. Mean ($\pm$ s.e.m.) blood plasma testosterone levels of nNOS^{-/-} and wild-type male mice before the start of the behavioural testing were 1.95 ± 0.22 and 2.14 ± 0.14 ng ml⁻¹, respectively ($P>0.05$); at the end of the experiment, mean ($\pm$ s.e.m.) testosterone values of nNOS^{-/-} and wild-type males were 1.86 ± 0.12 and 1.98 ± 0.01 ng ml⁻¹, respectively ($P>0.05$). The basal blood sample was obtained 10 days after the mice arrived in the behavioural laboratory and 2 weeks before the onset of behavioural testing. Another sample was obtained 1 h after the behavioural tests were completed. Blood samples were obtained from the retro-orbital sinus of mice that were lightly anaesthetized with methoxyflurane vapours (Metofane; Pitman Moore, Mundelein, IL). Blood was centrifuged at 3,500 r.p.m. for 1 h at 4 °C; plasma was separated and stored frozen (-80 °C) until assayed for testosterone. Blood plasma testosterone levels were assayed by radioimmunoassay (RIA) using an ¹²⁵I kit purchased from ICN Biochemicals (Carson, CA). The testosterone assay was highly specific; crossreactions with other steroid hormones were $<0.1\%$. The plasma testosterone values were determined in a single RIA. Because the s.e.m. exceeded the mean, a log-transform of the data was performed and statistical analysis of these transformed data were conducted. Many tests were used to assess coordinated behaviours, including orienting reactions, forelimb strength, coordination, climbing and locomotion. All sensorimotor tests were scored blindly. In one test of agility, the time required to turn in a blind alley (up to 2 min) was recorded. The mouse was placed facing the back wall of an alley (3 cm wide with walls 15 cm high). There was no difference between nNOS^{-/-} and wild-type mice in this task (22.0 ± 0.2



versus 22.3 ± 0.3 s, respectively; $P>0.05$). To assess locomotor balance and coordination, a mouse was placed at the centre of a wooden 'bridge' 60 cm long suspended 60 cm above a foam pillow. A pole bridge (2 cm in diameter) was used and the time taken for each mouse to reach the platform on either end of the bridge within 120 s was recorded. If the mouse fell, then the latency to fall within 120 s was recorded. There were no significant decreases in the performance of nNOS^{-/-} mice as compared to wild-type mice. To assess forelimb strength, a mouse was suspended by its forelimbs on a wire stretched between two posts 60 cm above a foam pillow. The time (s) until the mouse fell or 90 s had passed was recorded (a score of zero was assigned to animals that fell immediately; a score of 90 was given to mice that did not fall). There were two trials to this task. nNOS^{-/-} mice were not significantly impaired as compared to the performance of wild-type mice ($P>0.05$). Open field activity is a commonly used measure of anxiety levels. To assess open field activity, an animal was placed in an open area (1 m²) for 10 min. An observer recorded the movement of the mouse during the testing period. A border 4 cm from the wall was marked off and the space beyond the border was operationally defined as the open field. The amount of time spent in the open field was compared to the time the mouse was moving along the wall of the arena. There were no significant differences between nNOS^{-/-} and wild-type males (9.14 versus 8.99 min, respectively; $P>0.05$) in their open field behaviour.

Though NOS has been implicated in development of cultured PC12 cells¹⁸, iNOS rather than nNOS was primarily involved. If nNOS deletion markedly altered neuronal development, one would anticipate substantial neuroanatomical and gross psychomotor abnormalities in nNOS⁻ mice. However, detailed neuroanatomical studies⁹ and a thorough sensorimotor evaluation have not detected abnormalities. Also, synaptic plasticity has been evaluated in the nNOS⁻ mice revealing normal long-term potentiation in hippocampal slices¹⁰ and normal long-term depression in cerebellar cultures¹¹. Accordingly, it is highly probable that the behavioural abnormalities we have observed are direct, selective consequences of the loss of nNOS and not secondary to global physiological disruptions. Though mating and aggression are largely centrally mediated, a role for peripheral and indirect influences such as the dilated stomachs of nNOS⁻ mice cannot be completely excluded⁹. As drug-induced serotonin depletion augments aggressive and sexual activity^{22,23} and NO neurons contact serotonin cells⁵, the nNOS⁻ mice behavioural abnormalities might involve serotonin.

Our findings provide the first evidence for a behavioural role of central nNOS neurons, which presumably participate in sexual and aggressive behaviour. Studies administering NOS inhibitors have implicated NO in alcohol consumption¹⁹ and feeding behaviour²⁰. However, NOS inhibitors such as L-nitroarginine influence macrophage and endothelial NOS as well as nNOS and can affect other biological systems that use arginine. The use of mice with gene deletions overcomes such problems of nonspecificity.

Though direct comparisons are not feasible, the sexual and aggressive aberrations of nNOS⁻ mice seem more pronounced than those reported with deletion of other genes. Accordingly, NO may be a major mediator of sexual and aggressive behav-

iours, relevant for studies of their biological determination in humans as well as mice. Studies of monoamine oxidase A indicate the relevance to humans of aggressive behaviour in mice. Mice with deletions of the gene for this enzyme display excessive aggression²¹, and humans with low levels of the enzyme are also hyperaggressive²⁴. □

Received 6 July; accepted 11 October 1995.

1. Moncada, S. & Higgs, A. *New Engl. J. Med.* **329**, 2002–2012 (1993).
2. Ignarro, L. J. *Rev. Pharmacol. Toxicol.* **30**, 535–560 (1990).
3. Nathan, C. *FASEB J.* **6**, 3051–3064 (1992).
4. Nathan, C. & Xie, Q.-W. *Cell* **78**, 915–918 (1994).
5. Dawson, T. M. & Snyder, S. H. *J. Neurosci.* **14**, 5147–5159 (1994).
6. Brett, D. S. *et al. Neuron* **7**, 615–624 (1991).
7. Vincent, S. R. & Kimura, H. *Neuroscience* **4**, 755–784 (1992).
8. Rodrigo, J. *et al. Phil. Trans. R. Soc. B* **345**, 175–221 (1994).
9. Huang, P. L. *et al. Cell* **75**, 1273–1286 (1993).
10. O'Dell, T. J. *et al. Science* **265**, 542–546 (1994).
11. Linden, D. J. *et al. J. Neurosci.* **15**, 5098–5105 (1995).
12. Huang, Z. *et al. Science* **265**, 1883–1885 (1994).
13. Irikura, K. *et al. Proc. natn. Acad. Sci. U.S.A.* **92**, 6823–6827 (1995).
14. Brain, P. F. & Haug, M. *Psychoneuroendocrinology* **17**, 537–551 (1992).
15. Chen, C. *et al. Science* **266**, 291–294 (1994).
16. Saudou, F. *et al. Science* **265**, 1875–1878 (1994).
17. Ogawa, S. *et al. Endocr. Soc. Abstr.* (1995).
18. Peunova, N. & Enikolopov, G. *Nature* **375**, 68–73 (1995).
19. Rezuani, A. H. *et al. Pharmacol. Biochem. Behav.* **50**, 265–270 (1995).
20. Squadrito, F. *et al. Eur. J. Pharmacol.* **255**, 51–55 (1994).
21. Cases, O. *et al. Science* **268**, 1763–1766 (1995).
22. Albert, D. J. & Walsh, M. L. *Neurosci. Biobehav. Rev.* **8**, 5–24 (1984).
23. Salis, P. J. & Dewsbury, D. A. *Nature* **232**, 400–440 (1971).
24. Brunner, H. G. *et al. Science* **262**, 578–580 (1993).

ACKNOWLEDGEMENTS. We thank V. B. Mountcastle for valuable suggestions. Supported by USPHS grants, a Clinical Investigator Development Award, and an International Life Sciences Foundation grant to T.M.D.; American Foundation of AIDS Research, Alzheimer's Association, National Alliance for Research on Schizophrenia and Depression grants to V.L.D.; a USPHS grant, a contract and grants of the W. M. Keck Foundation and the Stanley Foundation Research Awards program to S.H.S.; USPHS grants to R.J.N. and to Hopkins Population Center (supported testosterone RIAS) and NIH and Bristol-Myers-Squibb grants to P.L.H. and M.C.F.

Multiple essential functions of neuregulin in development

Dirk Meyer & Carmen Birchmeier*

Robert-Roessle-Strasse 10, 13122 Berlin, Germany

NEUREGULIN (also called NDF, heregulin, GGF and ARIA) is a member of the EGF family which induces growth and differentiation of epithelial, glial and muscle cells in culture^{1–4}. The biological effects of the factor are mediated by tyrosine kinase receptors. Neuregulin can bind directly to erbB3 and erbB4 and receptor heterodimerization allows neuregulin-dependent activation of erbB2 (refs 1, 2, 5). A targeted mutation in mice reveals multiple essential roles of neuregulin in development. Here we show that neuregulin ^{-/-} embryos die during embryogenesis and display heart malformations. In addition, Schwann cell precursors and cranial ganglia fail to develop normally. The phenotype demonstrates that *in vivo* neuregulin acts locally and frequently in a paracrine manner. All cell types affected by the mutation express either erbB3 or erbB4, indicating that either of these tyrosine kinase receptors can be a component in recognition and transmission of essential neuregulin signals.

To analyse the function of neuregulin *in vivo*, we have mutated the neuregulin gene by homologous recombination in embryonic stem (ES) cells, and used these cells to create mice that carry the mutation. A variety of different protein isoforms are produced from the single neuregulin gene⁶. All contain an EGF-like domain sufficient for biological activity, of which α - and β -variants exist². In a first mutant allele, neuregulin ^{$\Delta 7$} , the neomycin resistance (*neo*) gene replaces exon 7, 8 and 9, which encode the carboxy-terminus of the EGF domain of all known variants

(Fig. 1a). In a second mutant allele, exon 6 is fused to β -galactosidase (neuregulin-LacZ). Exon 6 sequences downstream of the fusion point, which encode the amino terminus of the EGF-like domain of all known isoforms (Fig. 1b), are deleted. Healthy and fertile mice that carry either neuregulin mutation in a heterozygous state were generated. Matings between heterozygous animals produced no homozygous mutant offspring, indicating that such animals die during development. Genotyping of embryos revealed a mendelian ratio of homozygous mutant embryos up to and including day 10.5 in embryogenesis (E10.5). The survival of homozygous mutant embryos declined sharply, and none were alive on E11.5. They were found dead *in utero*, as judged by signs of resorption and absence of heartbeat. Analysis of mutant neuregulin locus (Fig. 1c) and mutant transcripts (Fig. 1d) demonstrate the presence of the desired mutation in the animals. Homozygous neuregulin-LacZ and neuregulin ^{$\Delta 7$} embryos display identical phenotypic changes.

Viable neuregulin ^{-/-} embryos on E10.5 were normal in overall size, but had irregular heartbeat and, frequently, an enlarged heart or pericard. Histological analysis showed that they had poorly developed ventricular trabeculae (Fig. 2a, b) but unchanged morphology of the atrium. Trabeculae are the first morphological sign of ventricle differentiation and are formed by heart muscle cells⁷. Because of the missing trabeculae, the ventricle and atrium had a similar morphology in neuregulin ^{-/-} embryos on E10.5, the endocardial cushion was not closed, and mesenchymal cells forming the cushion were reduced in number (Fig. 2a, b).

Expression analysis of neuregulin, erbB3 and erbB4 demonstrated the presence of all three molecules in the developing heart. X-Gal staining of neuregulin-LacZ/+ embryos, or *in situ* hybridization analysis with neuregulin-specific probes, demonstrated that neuregulin is produced in the endocardial endothelium (Fig. 2c, f), which does not express erbB3 or erbB4. Particularly strong neuregulin expression is associated with

* To whom correspondence should be addressed.

endothelia lining the endocardial cushion (arrowheads in Fig. 2c). The *erbB4* gene is expressed in the myocardium and myocardial cells that form the trabecules (Fig. 2d). Myocardial cells or endocardial endothelia do not express *erbB3* (Fig. 2e, g). Thus *erbB4* seems to participate in the control of myocardial cell outgrowth that ultimately results in trabeculation. In addition, transcripts for *erbB3* (Fig. 2e, g), but not *erbB4* (Fig. 2d), are detected in mesenchymal cells of the endocardial cushion and bulbus cordis. Thus *erbB3* seems to participate in development of the endocardial cushion, possibly in growth regulation of mesenchymal cells in the cushion. Local cellular interactions mediated by neuregulin and its receptors therefore regulate different steps in heart morphogenesis. Neuregulin is produced in all heart endothelia, but we observed clear differences in expression levels in endothelia lining endocardial cushion, ventricle or atrium (Fig. 2c). On a molecular level, the endothelia in different regions of the heart are thus different, although these cells show no overt distinction. Heart malformation is the most likely cause of death of neuregulin $-/-$ embryos.

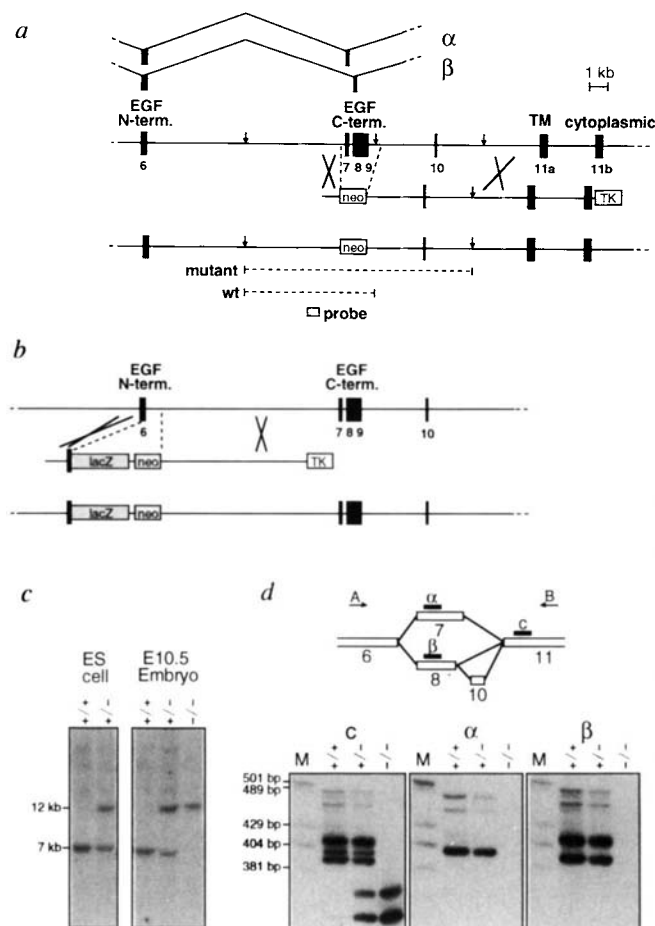
When patterns of cells expressing *erbB3* are compared in homozygous mutant and control littermates, two distinct changes in the developing peripheral nervous system become obvious: *erbB3*-expressing Schwann cell precursors and cranial ganglia cells are reduced in number. No change is apparent in dorsal root ganglia, myotomal or epithelial cells that also express *erbB3* (Fig. 3). In control E10.5 embryos, *erbB3*-expressing cells

are present along all peripheral projections (Fig. 3a–c, g, h; in Fig. 3c, h, arrowheads indicate *erbB3*-expressing cells along axons). Because of their location, these cells are classified as Schwann cell precursors. In neuregulin $-/-$ embryos, Schwann cell precursors in the trunk appear severely reduced in number, regardless of whether *erbB3* (Fig. 3d–f, i) or AP-2 probes were used for their identification. AP-2 encodes a transcription factor and is expressed in neural crest derivatives⁸. Neuregulin is expressed in peripheral sensory ganglia and motor neurons, that is, in nerves that will become ensheathed by Schwann cells^{3,4,9–11}. The reduced number might reflect a role of neuregulin in the modulation of cell fate decisions in the Schwann cell lineage, or in growth control of precursor cells already committed to a Schwann cell fate^{4,12}. We found that *erbB3* is expressed in developing Schwann cells and their precursors, the migrating neural crest, but we failed to detect *erbB4* by *in situ* hybridization in such cells. Thus *erbB3* seems to be responsible for neuregulin binding in developing glial cells, and might function together with *erbB2* as co-receptor; *erbB2* is present in Schwann cells^{12–14}.

Expression of *erbB3* was found in peripheral ganglia cells as well as Schwann cell precursors. The mutation affects *erbB3*-expressing cells in cranial or dorsal root ganglia to different extents. In cranial ganglia of neuregulin $-/-$ embryos, *erbB3*-positive cells within ganglia and along axonal projections are lost (Fig. 3a, b, d, e). This is in marked contrast to the trunk, where only *erbB3*-expressing Schwann cell precursors are affected (Fig. 3c, f).

FIG. 1 Outline of the strategy used to disrupt the neuregulin gene. **a**, Schematic representation of exon structures in the part of the neuregulin gene encoding the EGF domain (α or β variant). Targeting vector and mutant neuregulin gene are shown below. The open box (probe) corresponds to the sequence used for the Southern hybridization in **c**. The broken lines indicate the size of the *NdeI* fragment detected in wild-type and mutant DNA. Exon numbers are indicated. **b**, Schematic representation of wild-type neuregulin locus (top), targeting vector containing β -galactosidase (*LacZ*) fused to neuregulin coding sequences (middle), and mutant neuregulin–*LacZ* locus (bottom). **c**, Southern blot analysis of genomic DNA from wild-type (+/+), heterozygous (+/–) ES cells, and from wild-type (+/+), heterozygous (+/–) and homozygous (–/–) mutant embryos (E10.5). **d**, Strategy used for analysis of embryonic neuregulin transcripts by polymerase chain reaction (PCR). The schematic structure of alternatively spliced neuregulin transcripts (top) is shown; numbers indicate the exons present in different isoforms. For analysis of neuregulin transcripts from embryos, PCR analysis on complementary DNA synthesized from total embryonic poly(A)⁺ RNA was used. cDNA was amplified by PCR using primers A and B (indicated by arrows). Southern blot analyses (bottom) of PCR products with oligonucleotide α , β and γ (indicated by black boxes) as probes. cDNA from wild-type (+/+), heterozygous (+/–) and homozygous (–/–) mutant embryos was used for PCR.

METHODS. Genomic neuregulin DNA isolated from a library of 129 mouse DNA (isogenic to ES cell line E14) was used to construct a targeting vector. In this, exon 7, 8 and 9, encoding parts of the EGF domain of neuregulin (see ref. 3 for numbering of the exons) was replaced by the neomycin resistance gene (*neo*); the herpes simplex thymidine kinase (*tk*) gene²⁵ was fused at the 3' end. A second targeting vector containing β -galactosidase (*lacZ*) fused in frame to exon 6 of neuregulin was also generated. Targeting vectors were inserted into E14.1 (ref. 26) cells by electroporation; homologous recombination events were enriched by selection with G418 and gancyclovir and identified by Southern hybridization. The structure of the mutant loci and the presence of a single insertion event in ES cell colonies were verified by Southern hybridization. Two independently mutated ES clones carrying neuregulin mutation and one clone carrying the neuregulin–*lacZ* fusion were used to produce mutant mice by blastocyst injection²⁷. Phenotypes were analysed on C57BL/6J129 hybrid background. The analysis reported here is primarily of mice carrying the neuregulin^{A7–9} allele. For PCR analysis, poly(A)⁺ RNA was isolated from total embryonic RNA obtained from wild-type, heterozygous or homozygous mutant E10.5 embryos. cDNA was synthesized (Superscript, Gibco BRL), amplified with primers, and analysed by Southern hybridization and ³²P-endlabelled oligonucleotides as probes.



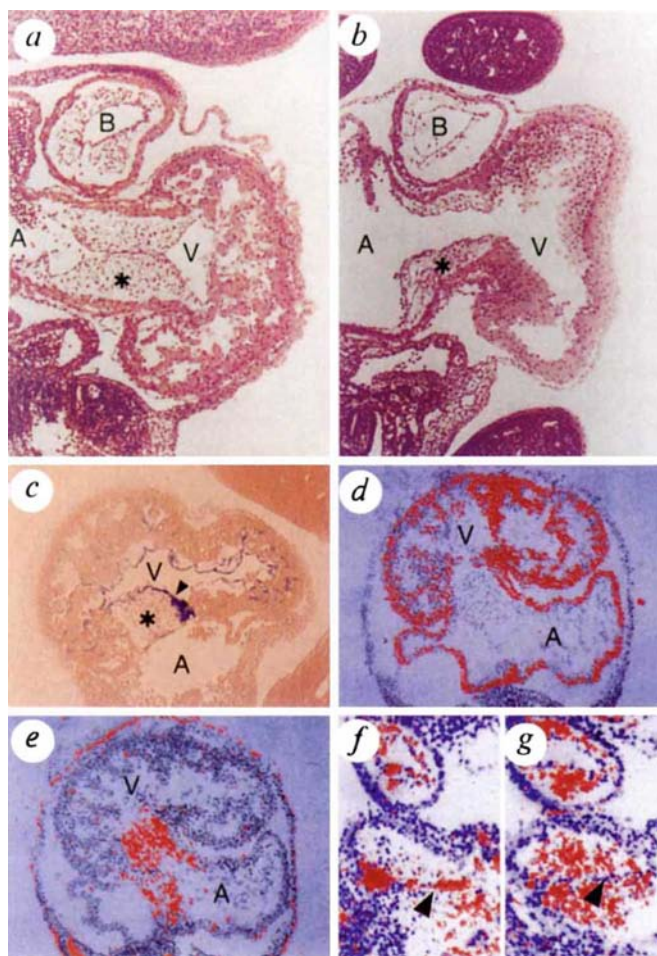


FIG. 2 Neuregulin is essential for normal heart morphogenesis during development. *a, b*, Haematoxylin- and eosin-stained sections of control neuregulin $+/+$ (*a*) and neuregulin $-/-$ (*b*) E10.5 embryos, demonstrating an absence of trabeculation in the ventricle (V) and distorted endocardial cushion (asterisks) in the neuregulin $-/-$ embryo; A, atrium; B, bulbus cordis. No damage to heart muscle or mesenchymal cells was apparent, and the endocard appeared normal upon histological and immunohistological analysis performed with endothelial-specific MECA32. *c*, Sections from neuregulin-lacZ/+ embryo stained with X-gal and counterstained with eosin, demonstrating expression of the neuregulin-LacZ fusion protein in endothelia lining the heart. Weak staining is associated with endothelia in the atrium (A); staining is more pronounced in endothelia of the ventricle (V). Arrow indicates the strongest-staining endothelia associated with the endocardial cushion (asterisks). *d, e*, Consecutive sections of E10.5 embryos analysed for expression of neuregulin receptors, erbB4 (*d*) and erbB3 (*e*) by *in situ* hybridization. *f, g*, Consecutive sections of E10.5 embryos analysed for expression of neuregulin (*f*) and erbB3 (*g*) by *in situ* hybridization. Note the presence of neuregulin, but not erbB3, in endothelia lining the endocardial cushion which are indicated by arrows.

METHODS. For haematoxylin and eosin staining, and X-gal²⁸ and eosin counter-staining, embryos were embedded in Technovit 7100 (Heraeus Kulzer); 8- μ m sections were used. For *in situ* hybridization analysis²⁹, frozen sections were hybridized to ³⁵S-labelled antisense mouse neuregulin, erbB3 or erbB4 RNA. False colour (pink) was used to display the hybridization signal observed in darkfield microscopy in *d-g*; the false-colour image was superimposed on a photograph of the Giemsa-stained section observed in brightfield. The corresponding sense controls did not show hybridization signals.

Immunohistochemistry performed with specific antibodies against neurofilament protein revealed a severely altered morphology of cranial nerves in neuregulin $-/-$ embryos, whereas other peripheral projections appeared unaffected (Fig. 4*a-d*). Distinct cranial ganglia were altered to different extents. Trigem-

FIG. 3 Analysis of Schwann cell precursors and peripheral ganglia in wild-type and neuregulin $-/-$ embryos. *In situ* hybridization analysis with erbB3-specific probes on neuregulin $+/+$ (*a-c*) and neuregulin $-/-$ E10.5 embryos (*d-f*). External appearance of the entire embryo (*a, d*); higher magnification of the head to show erbB3-expressing cells in cranial ganglia and otic vesicle (*b, e*); appearance of the trunk to show erbB3-expressing cells in developing dorsal root ganglia (*c, f*). Otic vesicle (ov), trigeminus (t), geniculate (g), petrose (p), nodose (n) and jugular (j) ganglia are indicated; erbB3-expressing cells aligned along axonal projections of the dorsal root ganglia or motor neurons are indicated by arrowheads. Section of neuregulin $-/+$ embryo (*g*) after *in situ* hybridization, demonstrating erbB3-expressing cells in the myotome (m), dorsal root ganglia (drg) and in Schwann cell precursors located along axonal projections of the motor neuron (mn) or sensory nerves (s). Higher magnification of sections obtained from neuregulin $-/+$ (*h*) or neuregulin $-/-$ embryos (*i*), demonstrating a reduced number of erbB3-expressing cells along motor neurons (indicated by arrow) and sensory nerves in mutant embryos. *In situ* hybridization on wild-type E9.25 embryo with neuregulin- (*j*), erbB3- (*k*) and erbB4- (*l*) specific probes, demonstrating that neuregulin and erbB3, but not erbB4, are expressed in neural crest cells emanating from the hindbrain. Note that in trunk neural crest, erbB3 is expressed but neuregulin or erbB4 are not (*k*). The arrows point towards rhombomers 2 and 4. Note that erbB4 is expressed in rhombomers 3 and 5, whereas neuregulin is expressed in a complementary fashion in the hindbrain, that is, in all rhombomers except 3 and 5.

METHODS. *In situ* hybridization on neuregulin $+/+$ and $-/-$ embryos was performed with antisense erbB3, erbB4 and neuregulin probes labelled by digoxigenin and detected with alkaline phosphatase conjugated anti-digoxigenin antibodies and 4-nitro-blue tetrazolium chloride²⁹ (Boehringer Mannheim RNA labelling and detection kit). After hybridization, entire embryos or frozen sections of the embryos were inspected.

inal nerves, particularly in the dorsal part, are lost (Fig. 4*c, d*). Mandibular projections and axonal connections between trigeminus and hindbrain are also absent (Fig. 4*c, d*). Geniculate and vestibular-cochlear ganglia appear little affected in mutant embryos. Petrose and nodose ganglia are reduced in size and their projections are disorganized; indeed petrosal projections towards the central nervous system are frequently not formed. The jugular ganglion is also abnormal (Fig. 4*c, d*). Uniform X-gal staining and histological appearance of all trigeminal cells of neuregulin-LacZ-neuregulin-LacZ embryos was apparent (Fig. 4*f*), whereas the trigeminus of neuregulin-LacZ/+ embryos appeared heterogeneous (Fig. 4*e*). Two cell populations, neural crest cells from the hindbrain and ectodermal cells from placodes, participate in the development of cranial ganglia^{15,16}. The ganglion structures least affected by the neuregulin mutation are mostly placodal in origin¹⁵. Ablation of neural crest cells in the chick leads to changes in cranial ganglia that are similar, if not identical, to the changes observed in neuregulin $-/-$ mouse embryos¹⁷. We suggest, therefore, that neural crest cells do not contribute to cranial ganglia in neuregulin $-/-$ embryos, resulting in simultaneous loss of crest-derived Schwann cells and neurons.

To investigate further the cause of the development aberrance of cranial ganglia, we analysed the expression of neuregulin, erbB3 and erbB4 during cranial ganglia development, and found complex expression patterns. Neuregulin is expressed in the hindbrain (rhombomers 2, 4 and 6) and in migrating cranial cells, but not in trunk neural crest cells (Fig. 3*j*). ErbB3 is expressed in migrating neural crest cells of the head and trunk, but not in the central nervous system (Fig. 3*k*), and erbB4 is

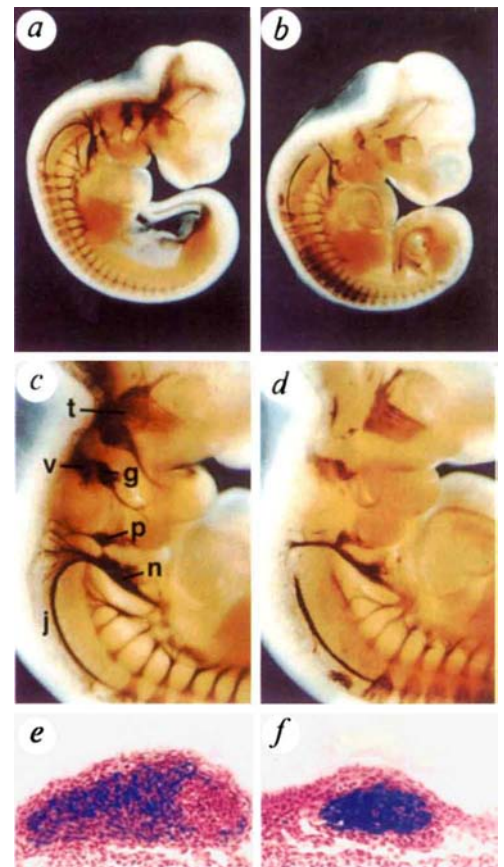
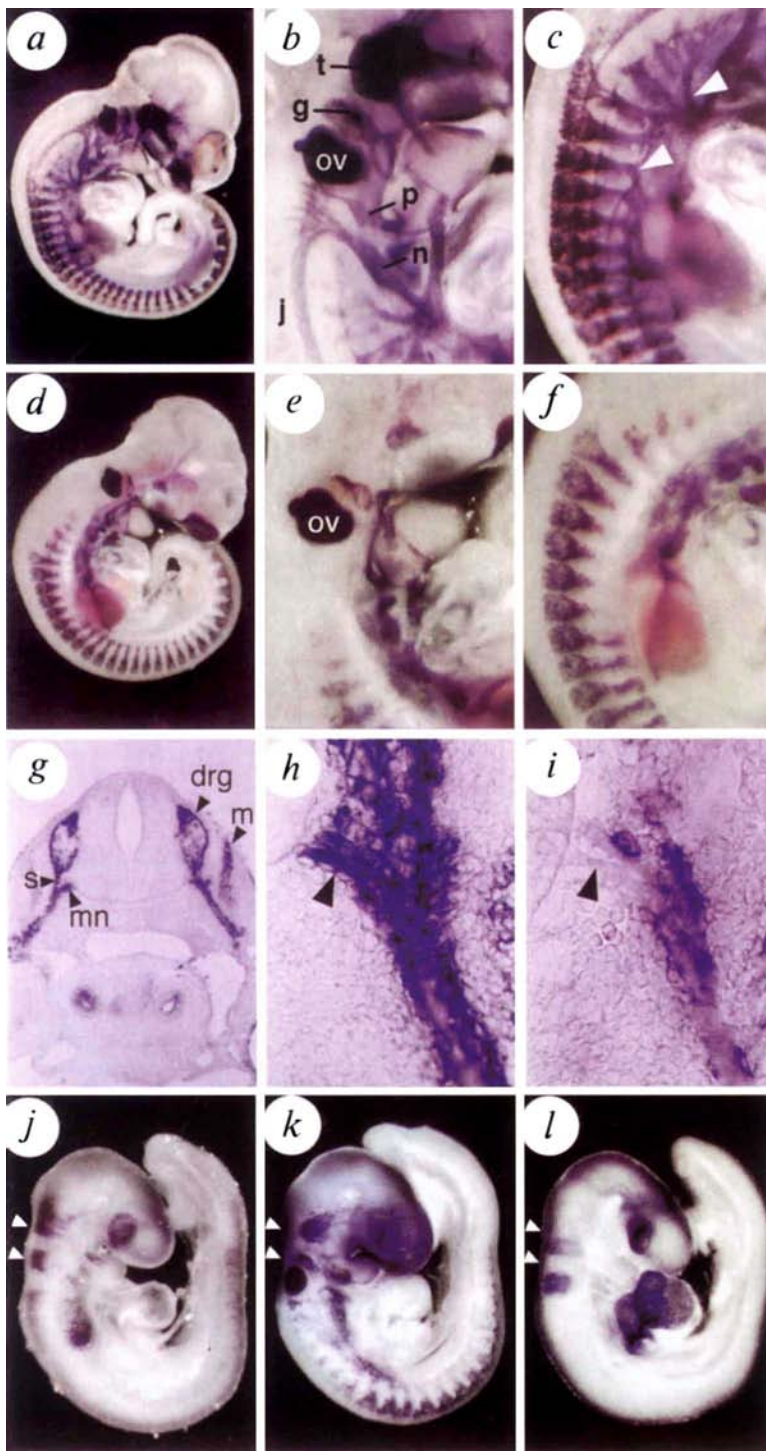


FIG. 4 Cranial ganglia are morphologically aberrant in neuregulin $-/-$ embryos. Nerves of neuregulin $-/+$ (a, c) or neuregulin $-/-$ (b, d) E10.5 embryos were visualized by immunohistochemistry using anti-NF160 specific antibodies. Note that at this stage in development, NF160 is expressed mainly in the peripheral nervous system. Cranial ganglia (trigeminal, t; geniculate, g; vestibulocochlear, v; petrosal, p; and jugular, j) are indicated. Sections of neuregulin-LacZ/+ (e) or neuregulin-LacZ/neuregulin-LacZ (f) embryo stained with X-gal and counterstained with haematoxylin eosin. In the trigeminal of neuregulin-LacZ/+ embryos (e), β -gal-staining cells are intermingled with non-staining cells and heterogeneous cellular body sizes are observed; in contrast, all trigeminal cells from neuregulin-LacZ/neuregulin-LacZ embryos (f) displayed positive β -gal staining and large cell bodies.

METHODS. Embryos were fixed in 5% acidic ethanol and neurofilament was visualized with anti-NF160 antibodies (Sigma). Sections were stained with X-gal and counterstained with eosin and haematoxylin. Note that X-gal staining is not observed in all neuregulin-expressing cells (see refs 3, 4, 9–11); only neuregulin isoforms that do not provide a signal peptide to the fusion protein are expected to give a positive signal³⁰.

expressed in the hindbrain (rhombomers 3 and 5) but not in the peripheral nervous system (Fig. 3f). The segmental expression of Krox 20 (ref. 18), HoxB1 (ref. 19), neuregulin and erbB4 are not altered, so hindbrain segmentation is not affected in the mutant embryos. Migrating neural crest cells labelled by various markers (Krox 20 (refs 18, 20), HoxB1 (ref. 19) and AP-2 (ref. 8)) appear similar in mutant and control littermates, indicating that such cells are formed and their migration is not affected. However, the pattern of erbB3-expressing cells in neuregulin $-/-$ embryos and control littermates was altered when cranial ganglia started to form, and the number of erbB3-expressing cells in cranial structures is already reduced on E9.25. In the

cranial neural-crest lineage, neuregulin affects the neurogenic subpopulation and mesenchymal derivatives appear not to be altered at the developmental stages analysed. Thus neuregulin might function as a survival factor for neurogenic cells of the cranial neural crest. Alternatively, the factor might allow cranial crest cells to assume a neurogenic fate or enter the ganglia. The expression patterns of neuregulin in cranial and trunk neural crest cells are distinct in that neuregulin is expressed only in cranial, but not trunk, neural crest cells during migration. Our results emphasize the disparity between neurogenic neural crest of the head and of the trunk²¹: in cranial neural crest, neuregulin is essential for the neurogenic lineage, whereas in trunk neural

crest, the factor affects early steps in Schwann cell development.

Mutation of the neuregulin gene in the mouse has indicated that it has multiple independent and essential functions in development. The distinct expression patterns of neuregulin and its direct receptors, erbB3 and erbB4, had previously indicated that this signalling system can function in a paracrine manner. Indeed, cells lacking functional neuregulin do not usually express the factor but are located close to neuregulin-expressing cells. This indicates a local and paracrine signalling mode for neuregulin *in vivo*. Several neuregulin isoforms are known to bind to heparin, a characteristic used previously for purification^{1,4}. Heparin sulphate, which is present in high concentrations on cell surfaces, might interfere with free diffusion and long-range signalling of neuregulin. All cell types that are affected by the neuregulin mutation express either erbB3 or erbB4. In particular, multiple phenotypic changes are found in cells that express erbB3; such cells might require a co-receptor for receiving and transmitting neuregulin signals because erbB3 might not have catalytic protein kinase activity. EGF receptor and erbB2 can function as co-receptors *in vitro*. Mutation of the EGF receptor has indicated a phenotype distinct from that described here^{22,24}. Genetic analysis of the erbB2, erbB3 and erbB4 receptors will lead to further understanding of this complex signalling system. □

Received 21 July; accepted 3 October 1995.

1. Wen, D. et al. *Cell* **69**, 559–572 (1992).
2. Holmes, W. E. et al. *Science* **256**, 1205–1210 (1992).
3. Falls, D. L., Rosen, K. M., Corfas, G., Lane, W. S. & Fischbach, G. D. *Cell* **72**, 801–815 (1993).
4. Marchionni, M. A., et al. *Nature* **362**, 312–318 (1993).
5. Carraway, K. L. & Cantley, L. C. *Cell* **78**, 5–8 (1994).

6. Peles, E. & Yarden, Y. *Bioessays* **15**, 815–824 (1993).
7. Icardo, J. M. & Fernandez, T. A. *Acta anat.* **130**, 264–274 (1987).
8. Mitchell, P. J., Timmons, P. M., Hebert, J. M., Rigby, P. W. & Tjian, R. *Genes Dev.* **5**, 105–119 (1991).
9. Meyer, D. & Birchmeier, C. *Proc. natn. Acad. Sci. U.S.A.* **91**, 1064–1068 (1994).
10. Corfas, G., Rosen, K. M., Aratake, H., Krauss, R. & Fischbach, G. D. *Neuron* **14**, 103–115 (1995).
11. Ho, W. H., Armanini, M. P., Nuijens, A., Phillips, H. S. & Osheroff, P. L. *J. Biol. Chem.* **270**, 14523–14532 (1995).
12. Shah, N. M., Marchionni, M. A., Isaacs, I., Stroobant, P. & Anderson, D. J. *Cell* **77**, 349–360 (1994).
13. Jin, J. J., Nikitin, A. & Rajewsky, M. F. *Cell Growth Differ.* **4**, 227–237 (1993).
14. Morrissey, T. K., Levi, A. D., Nuijens, A., Sliwkowski, M. X. & Bunge, R. P. *Proc. natn. Acad. Sci. U.S.A.* **92**, 1431–1435 (1995).
15. D'Amico, M. A. & Noden, D. M. *Am. J. Anat.* **166**, 445–468 (1983).
16. LeDouran, N. M., Fontaine-Perus, J. & Couly, G. *Trends Neurosci.* **9**, 175–180 (1986).
17. Hamburger, V. *J. exp. Zool.* **148**, 91–124 (1961).
18. Wilkinson, D. G., Bhatt, S., Chavrier, P., Bravo, R. & Charnay, P. *Nature* **337**, 461–464 (1989).
19. Murphy, P., Davidson, D. R. & Hill, R. E. *Nature* **341**, 156–159 (1989).
20. Sham, M. H. et al. *Cell* **72**, 183–196 (1993).
21. Bronner-Fraser, M. *Expl. Cell Res.* **118**, 405–417 (1995).
22. Threadgill, D. W. et al. *Science* **269**, 230–234 (1995).
23. Sibilia, M. & Wagner, E. F. *Science* **269**, 234–238 (1995).
24. Miettinen, P. J. et al. *Nature* **376**, 337–341 (1995).
25. Mansour, S. L., Thomas, K. R. & Capecchi, M. R. *Nature* **336**, 348–352 (1988).
26. Kuhn, R., Rajewsky, K. & Muller, W. *Science* **254**, 707–710 (1991).
27. Bradley, A. in *Tetrahymena and Embryonic Stem Cells: A Practical Approach* (ed. Robertson, E.) 113–157 (IRL, Oxford, 1987).
28. Gossler, A. & Zachgo, J. in *Gene Targeting: A Practical Approach* (ed. Joyner, A. L.) 207–215 (IRL, New York, 1993).
29. Wilkinson, D. G. in *In Situ Hybridization: A Practical Approach* (Oxford Univ. Press, 1992).
30. Skarnes, W. C., Moss, J. E., Hurlley, S. M. & Beddington, R. *Proc. natn. Acad. Sci. U.S.A.* **92**, 6592–6596 (1995).

ACKNOWLEDGEMENTS. Much of this work was done at the Max-Planck-Laboratorium in der Max-Planck-Gesellschaft, Köln. We thank W. Birchmeier, D. Wilkinson, A. Aguzzi, S. Isenmann and A. Davies for discussions and critical reading of the manuscript; B. Bettenhausen for introducing us to whole-mount *in situ* hybridizations; D. Wilkinson (HoxB1, Krox20) and P. Mitchell (Ap-2) for cDNA clones used for *in situ* hybridization; Y. Yarden for human cDNAs for erbB3 and erbB4; R. Hallmann for endothelial-specific monoclonal antibodies; A. Rehaus for technical assistance; and B. Lemke for help with computer programs used for figure preparation. This study was supported by grants from the Deutsche Forschungsgesellschaft and the Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie to C.B.

Aberrant neural and cardiac development in mice lacking the ErbB4 neuregulin receptor

Martin Gassmann*, Franca Casagrande†, Donata Orioli†, Horst Simon*, Cary Lai‡, Rüdiger Klein† & Greg Lemke*§

* Molecular Neurobiology Laboratory, Salk Institute for Biological Studies, La Jolla, California 92037, USA

† Differentiation Program, European Molecular Biology Laboratory, 69012 Heidelberg, Germany

‡ Department of Neuropharmacology, Scripps Research Institute, La Jolla, California 92037, USA

VARIOUS *in vitro* studies have suggested that ErbB4 (HER4) is a receptor for the neuregulins, a family of closely related proteins implicated as regulators of neural and muscle development, and of the differentiation and oncogenic transformation of mammary epithelia^{1–3}. Here we demonstrate that ErbB4 is an essential *in vivo* regulator of both cardiac muscle differentiation and axon guidance in the central nervous system (CNS). Mice lacking ErbB4 die during mid-embryogenesis from the aborted development of myocardial trabeculae in the heart ventricle. They also display striking alterations in innervation of the hindbrain in the CNS that are consistent with the restricted expression of the *ErbB4* gene in rhombomeres 3 and 5. Similarities in the cardiac phenotype of *ErbB4* and *neuregulin* gene mutants suggest that ErbB4 functions as a neuregulin receptor in the heart; however, differences in the hindbrain phenotypes of these mutants are consistent with the action of a new ErbB4 ligand in the CNS.

ErbB4 (HER4), ErbB3 (HER3) and ErbB2 (HER2, Neu) are a family of cell-surface receptors that exhibit structural similarity to the receptor for epidermal growth factor (EGF)^{4,6}. Although these proteins have been intensively studied in a variety of biological contexts in cell culture, their action and interaction during mammalian development *in vivo* are largely unknown. In cultured cells, the ligands for the ErbB2/3/4 receptors are the neuregulins^{1,7}, a set of proteins also referred to as glial growth factors (GGFs)^{8,9}, Neu differentiation factors (NDFs)¹⁰, heregulins (HRGs)¹¹ and acetylcholine receptor inducing activity (ARIA)¹². This surfeit of names reflects the diverse biological activities of the neuregulins *in vitro*, as glial cell mitogens, receptor binding proteins, mammary differentiation factors, and muscle trophic factors.

We sought to assess the *in vivo* action of the most functionally independent of the neuregulin receptors—ErbB4 (ref. 1)—through the generation of loss-of-function mutations in the mouse *ErbB4* gene. We isolated genomic clones of this gene from a strain 129/Sv genomic library, prepared the pPNT-based targeting vector (Fig. 1a)¹³, electroporated this construct into the R1 line of mouse embryonic stem (ES) cells¹⁴, and identified clonal lines in which the mutated *ErbB4* allele had replaced one of the wild-type *ErbB4* alleles through homologous recombination (Fig. 1b). Selected ES cell clones were microinjected into C57BL/6J blastocysts, which were then transferred to pseudo-pregnant females to generate chimaeric mice. Breeding of these chimaeras to C57BL/6J mice resulted in germline transmission of the *ErbB4* mutation. Mice heterozygous for the inactivated *ErbB4* allele displayed no obvious behavioural or anatomical defects. When these heterozygotes were crossed, however, we observed no homozygous mutants in any of the newborn litters examined. Closer inspection showed that all *ErbB4*^{−/−} homozygotes died *in utero*, between 10 and 11 days after fertilization (E10–11). As expected from the design of the targeting vector (see Fig. 1), these homozygotes expressed no detectable ErbB4

§ To whom correspondence should be addressed.

protein, as assessed by western blot with an ErbB4 antibody (Fig. 1c) or *ErbB4* messenger RNA, as assessed by whole-mount *in situ* hybridization (see below).

The mid-embryonic lethality of *ErbB4*^{-/-} mice results from the aborted development of heart muscle. This phenotype is consistent with the normal pattern of *ErbB4* gene expression at E9.5, which is largely confined to cardiac muscle and the nervous system (Fig. 2). Within the heart, *ErbB4* mRNA is present throughout both the atrial and ventricular myocardium (muscle), but is absent from the endocardium, which is the endo-

thelial lining of the organ (Fig. 2c). (These endothelial cells are the site of *neuregulin* gene expression¹⁵.) *ErbB4* mRNA is present in both the myocardium of the developing chamber walls and in the myocardial cells that make up the trabeculae¹⁶, which are the interlinked finger-like projections of heart muscle that extend from the ventricular wall into the interior of the chamber (Figs 2d and 3a). We found that the E10.5 hearts of *ErbB4*^{-/-} mice contain intact, apparently normal, chamber walls but are devoid of ventricular trabeculae (Fig. 3a, b). The endocardium appears to be normally configured in the mutants, although the endo-

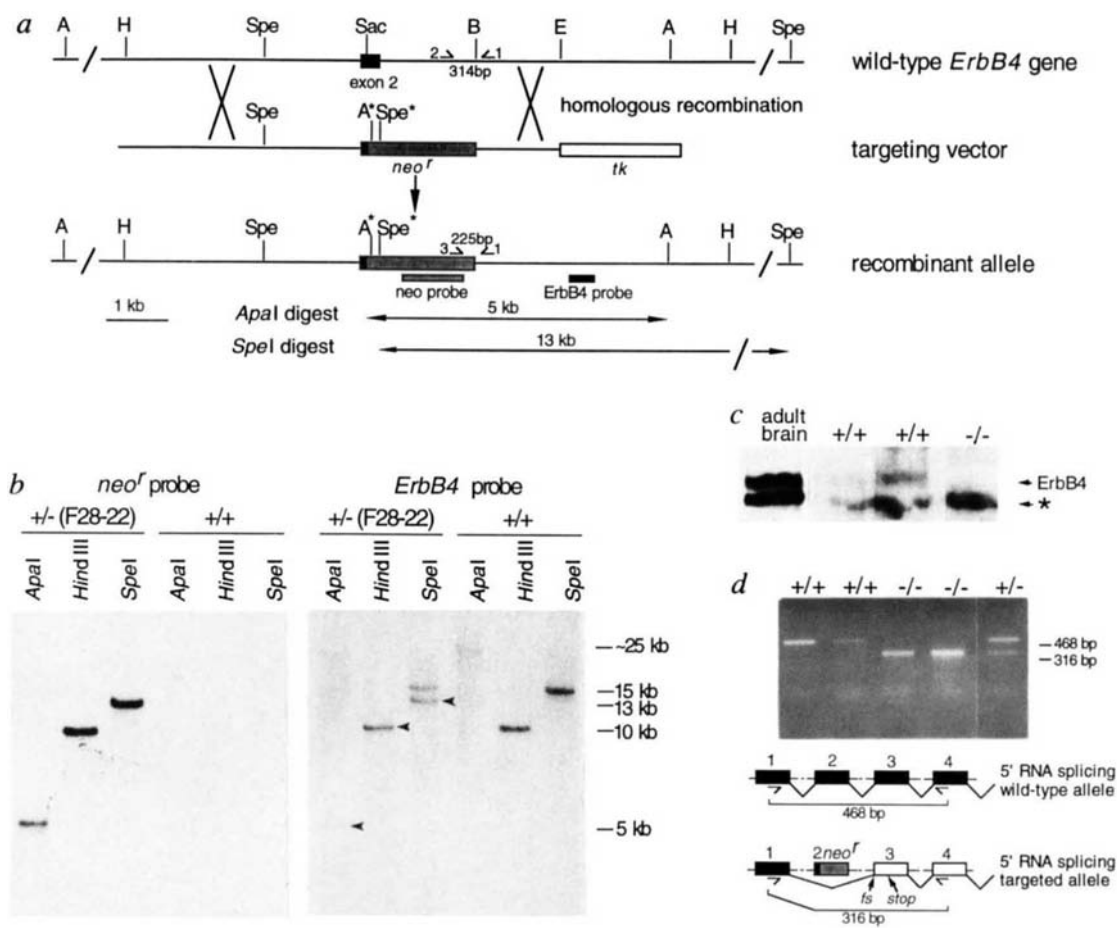


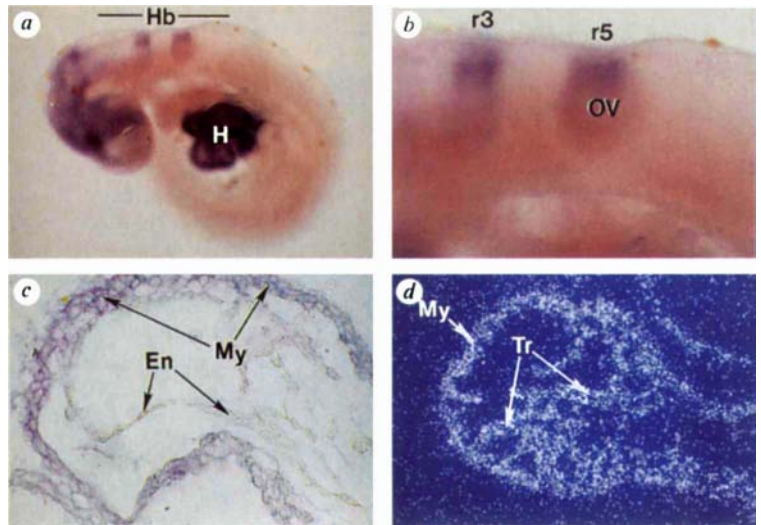
FIG. 1 Mouse *ErbB4* gene targeting. *a*, Structure of the mouse *ErbB4* gene around exon 2, of the targeting construct for homologous recombination, and of the recombinant allele. The *ErbB4* gene is >100 kb in length, and the λ phage clone used for preparation of the targeting construct contains only coding sequence from the second exon of the gene (M. Gassmann, unpublished data). Abbreviations: A, *Apal*; H, *HindIII*; Spe, *SpeI*; Sac, *SacI*; B, *BamHI*; E, *EcoRI*; *neo^r*, G418 resistance cassette for positive selection of homologous ES cell recombinants; *tk*, herpes simplex virus thymidine kinase gene for selection against non-homologous recombinants. The *neo* and *ErbB4* probes used for the Southern blots of *b*, and the polymerase chain reaction (PCR) primers (arrows 1, 2 and 3) used for routine genotyping of embryos, are indicated. New *Apal* and *SpeI* sites introduced through homologous recombination are indicated by asterisks. *b*, Southern blots of genomic DNA from wild-type ES cells (+/+) and from F28-22 ES cells heterozygous for the recombinant *ErbB4* allele (+/-), hybridized with either the *neo^r* (left) or *ErbB4* (right) probes indicated in *a*. Arrowheads in the *ErbB4* blot indicate hybridizing bands corresponding to the recombinant allele. The arrowed *HindIII* band of +/- DNA in the *ErbB4* blot is a doublet. *c*, Western blot of proteins extracted from adult mouse brain, and from wild-type (+/+) and *ErbB4* mutant (-/-) E10.5 mouse embryos, probed with an anti-ErbB4 rabbit antiserum (C. Lai, unpublished results). The p180 ErbB4 band and a crossreacting non-ErbB4 band (asterisk) are indicated. *d*, Top, reverse transcription (RT)-PCR of RNA extracted

from whole E10.5 wild-type (+/+), heterozygous (+/-) and *ErbB4* mutant (-/-) mouse embryos. *ErbB4*^{-/-} embryos contain a low level of mis-spliced, frameshifted *ErbB4* RNA that is not detectable by *in situ* hybridization. Bottom, diagram of deduced 5' splicing reactions, based on cloning and DNA sequencing of the 468-bp and 316-bp PCR products illustrated above, and comparison of these sequences to the previously mapped splice junctions flanking *ErbB4* exons 1-4 (M. Gassmann, unpublished results). Arrows indicate the position of the exon 1 and exon 4 primers used for RT-PCR; fs, frameshift.

METHODS. Southern blot hybridizations were performed according to standard protocols. Blots were hybridized with the indicated ³²P-dCTP-labelled probes. Western blots were performed according to standard protocols, and were reacted with an affinity-purified rabbit antiserum generated against a glutathione-S-transferase fusion protein containing carboxy-terminal-proximal amino acids of the mouse ErbB4 protein corresponding to residues 1185-1238 of the human ErbB4 sequence (C. Lai, unpublished data). Blots were developed with an ECL chemiluminescent detection system. RT-PCR was performed according to standard protocols, starting with ~5 μ g of total RNA isolated from whole mouse embryos. PCR was performed for 30 cycles, using upstream and downstream primers (arrows) corresponding to positions (-12)-12 and 456-433, respectively, from the start codon of the mouse *ErbB4* cDNA (C. Lai, unpublished results). Amplified DNAs were subcloned into pCR-Script (Stratagene) and sequenced.

FIG. 2 Expression of the wild-type *ErbB4* gene in E9.5 mouse embryos. **a**, Low-power view (anterior to the left) of an entire E9.5 embryo in which *ErbB4* mRNA (dark purple reaction product) is visualized with digoxigenin-based whole-mount *in situ* hybridization. *ErbB4* mRNA is evident in the heart (H) and rostral CNS, with two stripes of expression in segments of the hindbrain (Hb). **b**, Higher magnification view of the E9.5 hindbrain. *ErbB4* mRNA is confined to the dorsal regions of rhombomeres 3 and 5 (r3 and r5), the latter of which is positioned adjacent to the otic vesicle (OV). **c**, Sagittal section of an E9.5 heart (ventricle) from the embryo previously analysed by whole-mount *in situ* hybridization as in **a**. Ventricular trabeculation (see text) begins around this time. *ErbB4* mRNA is detectable in the myocardium (My), but not in the endocardium (En) that lines the ventricle. **d**, Sagittal section of a slightly older (E10.5) heart ventricle in which the formation of trabeculae (Tr) is well advanced. *ErbB4* mRNA, detected as white grains with ^{35}S *in situ* hybridization, is now apparent in both the myocardial wall (My) and the developing myocardial trabeculae.

METHODS. Whole-mount *in situ* hybridization with digoxigenin-labelled probes and detection with AP-conjugated antibody (Boehringer-Mannheim) was performed as described previously²⁸. After visualization, embryos were embedded in paraffin and sections were cut at 8 μm . For ^{35}S *in situ* hybridization embryos were fixed at 4% paraformaldehyde, immersed in 20% sucrose/PBS for 24 h and were subsequently embedded in OCT (Tissue-Tek). Hybridization to 14- μm cryosections was performed as described



previously^{29,30}, using ^{35}S -radiolabelled antisense and sense (control) RNAs transcribed *in situ* from a 5' mouse *ErbB4* cDNA fragment subcloned into pCR-Script.

cardial cushion, from which the heart valves eventually develop, is often slightly reduced in size relative to wild type (Fig. 3b). The failure of trabeculation in *ErbB4*^{-/-} mice leads to severely reduced embryonic blood flow: note the greater number of red blood cells present in the mutant heart (Fig. 3b) relative to wild type (Fig. 3a). The E10.5 embryonic lethality and lack of cardiac trabeculation that we observe in *ErbB4*^{-/-} mice are also prominent features of mice homozygous for loss-of-function mutations in either the *neuregulin* or *ErbB2* genes (see refs 17, 18).

In the E9.5 nervous system, *ErbB4* expression is observed in the hindbrain (Fig. 2b), the midbrain and the ventral forebrain

(Fig. 2a). Expression in the E9.5 hindbrain is precisely localized to two segments—rhombomeres 3 and 5—and is further restricted within these rhombomeres to the dorsal third of the neural tube (Fig. 2b) and to two symmetric pools of cells immediately above the developing ventral motor neurons (data not shown). The flanking segments that lack *ErbB4* mRNA—rhombomeres 2, 4 and 6—are associated with streams of neural crest cells that populate cranial sensory ganglia V (trigeminal), VII and VIII (facial and acoustic), and IX (glossopharyngeal) of the peripheral nervous system (PNS), respectively^{19,20}. We examined the organization and innervation of these ganglia in wild-type

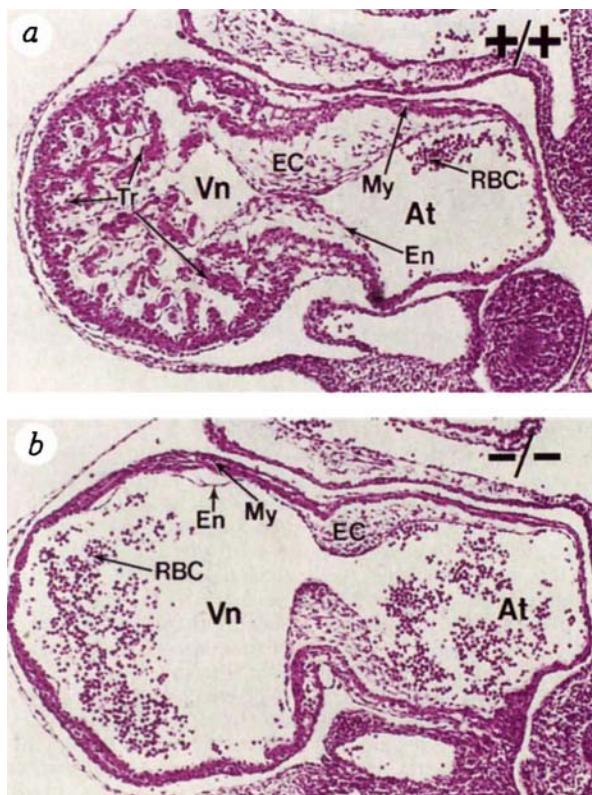


FIG. 3 Histology of wild-type and mutant mouse hearts at embryonic day 10.5. **a**, A sagittal section through the centre of wild-type (+/+) heart, in which the ventricle (Vn) and atrium (At) are separated by the endocardial cushion (EC). The outer myocardial (My) muscle and inner endocardial (En) endothelial tissues are readily distinguished, and elaborate myocardial trabeculae (Tr), draped by a layer of endocardial cells, are apparent in the ventricle. The atrium contains a small pool of red blood cells (RBCs). **b**, An equivalent sagittal section through the same region of an *ErbB4* mutant (-/-) heart. The mutant resembles the wild-type, except for the complete absence of ventricular trabeculae, the appearance of a slightly reduced endocardial cushion, and a noticeable swelling of both the atrium and ventricle. Many more RBCs have collected in the mutant heart chambers relative to wild-type, which is a typical feature of the poorly functioning mutant hearts.

METHODS. E10.5 wild-type and mutant mouse embryos were fixed in 4% paraformaldehyde and subsequently embedded in paraffin. Sections were cut at 8 μm , and were stained with haematoxylin-eosin according to standard protocols.

and mutant embryos at E10.5 by using an antibody (TuJ1) directed against a neuron-specific tubulin isoform²¹. These analyses indicated a near-fusion of ganglia V and VII/VIII. In addition in some but not all mutants, we observed a caudal displacement of cranial nerve IX towards nerve X (vagal) (Fig. 4a, b). Both of these alterations are also observed in mice carrying mutations in the *Krox-20* gene^{22,23}, which encodes a zinc-finger transcription factor which has expression restricted to r3 and r5 (ref. 24). (The caudal displacement of the IXth nerve, which may be due to the physical interference of the r6 neural-crest cell stream by a caudally displaced otic vesicle, is more robust in *Krox-20*^{-/-} mice than in *ErbB4* mutants^{22,23}.) Although these superficial similarities in the *Krox-20* and *ErbB4* mutant phenotypes are suggestive, *ErbB4* does not appear to regulate expression of the *Krox-20* gene, because (1) *Krox-20* mRNA is still detected in r3 and r5 in the *ErbB4* mutants (data not shown), (2) *Krox-20* is normally expressed more broadly along the dorsal-ventral axis of the neural tube than is *ErbB4*, and (3) r3 and r5 are largely intact at E10.5 in *ErbB4*^{-/-} mice, but are almost entirely deleted by this time in *Krox-20*^{-/-} mice^{22,23}. That *Krox-20* may participate in the rostral caudal regulation of the *ErbB4* gene remains a possibility.

The near-fusion of ganglia V and VII/VIII in *ErbB4*^{-/-} mice is not, as in the *Krox-20* mutants, due to the loss of r3 and r5, but rather to aberrant innervation from and to the hindbrain. Normally, motor axons destined for the trigeminal ganglion exit, and PNS sensory axons from the trigeminal enter, the hindbrain at r2, whereas axons to and from the facial/acoustic ganglia enter and exit the hindbrain at r4 (Fig. 4c)²⁵. In the *ErbB4* mutants, however, the trigeminal ganglion displays connections to the CNS at both r2 and r4 (Fig. 4d), and the facial and acoustic ganglia are also connected to the CNS at these same two rhombomeres (Fig. 4e). In both cases, the number of axons making up the incorrect connection is approximately the same as the number of axons in the correct connection. Because the loss of *ErbB4* in r3 and r5 leads to aberrant innervation at r2 and r4, we investigated whether in *ErbB4*^{-/-} hindbrains r2 had acquired molecular properties of r4, or vice versa. We have found no evidence for this: markers that distinguish r2 from r4 remain appropriately restricted in their expression in *ErbB4* mutants. For example, the r4-specific homeobox gene *HoxB1* (ref. 26) is neither lost from r4 (as predicted for an r4 to r2 transformation) nor acquired by r2 (Fig. 4f). To investigate hindbrain innervation in *ErbB4* mutants in greater detail, we performed local injections of the lipophilic tracer DiI into either the trigeminal or facial ganglion of E10.5 wild-type and *ErbB4*^{-/-} embryos. As expected, injection into the wild-type trigeminal ganglion labelled a single sensory axon projection to r2 and motor neuron cell bodies in r3, r2 and r1, whereas injection into the wild-type facial ganglion labelled a single sensory axon projection to r4 and motor neuron cell bodies in r4 and r5 (Fig. 4g). In contrast, local injection of DiI into the *ErbB4*^{-/-} facial ganglion labelled two sensory axon entry points in r4 and r2, as well as motor neuron cell bodies and axons in r4 and r5 (their normal location) and in r1, r2 and r3 (Fig. 4h). For the facial ganglion, equal numbers of sensory axons project to the correct (r4) and incorrect (r2) targets, and equal numbers of motor neuron cell bodies are retrogradely labelled in the correct and incorrect hindbrain segments. For the aberrant connection of the trigeminal ganglion to r4 (Fig. 4h), the primary defect appears to be in sensory innervation from the PNS.

We favour the hypothesis that the mis-targeting of axons to and from r2- and r4-derived cranial ganglia in the *ErbB4*^{-/-} hindbrain results from the loss of an *ErbB4*-dependent barrier molecule, expressed in r3 and r5, which is inhibitory to axon growth across these rhombomeres. It is possible that this barrier is *ErbB4* itself, or alternatively, that *ErbB4* activation is required for the expression of the barrier (perhaps a ligand for an Eph-related receptor). Neither *neuregulin* nor *ErbB2* gene mutants exhibit the hindbrain mis-innervation phenotype of the

ErbB4^{-/-} mice. The most parsimonious interpretation of this difference is that *ErbB4* recognizes a distinct ligand in this part of the CNS. A second *neuregulin*-related gene has been identified and cloned²⁷, and it is now of interest to determine whether this gene encodes a new *ErbB4* ligand. □

Received 16 August; accepted 6 October 1995.

1. Plowman, G. D. et al. *Nature* **366**, 473–475 (1993).
2. Marchionni, M. A. et al. *Nature* **362**, 312–318 (1993).
3. Carraway, K. L. R. & Burden, S. J. *Curr. Opin. Neurobiol.* **5**, 1–7 (1995).
4. Schechter, A. L. et al. *Nature* **312**, 513–516 (1984).
5. Kraus, M. H., Issing, W., Miki, T., Popescu, N. C. & Aaronson, S. A. *Proc. natn. Acad. Sci. U.S.A.* **86**, 9193–9197 (1989).
6. Plowman, G. D. et al. *Proc. natn. Acad. Sci. U.S.A.* **90**, 1746–1750 (1993).
7. Sliwkowski, M. X. et al. *J. Biol. Chem.* **269**, 14661–14665 (1994).
8. Lemke, G. E. & Brookes, J. P. J. *Neurosci.* **4**, 75–83 (1984).
9. Goodearl, A. D. et al. *J. Biol. Chem.* **268**, 18095–18102 (1993).
10. Wen, D. et al. *Cell* **69**, 559–572 (1992).
11. Holmes, W. E. et al. *Science* **256**, 1205–1210 (1992).
12. Falls, D. L., Rosen, K. M., Corfas, G., Lane, W. S. & Fischbach, G. D. *Cell* **72**, 801–815 (1993).
13. Tybulewicz, V. L., Crawford, C. E., Jackson, P. K., Bronson, R. T. & Mulligan, R. C. *Cell* **65**, 1153–1163 (1991).
14. Nagy, A., Rossant, J., Nagy, R., Abramow-Newerly, W. & Roder, J. C. *Proc. natn. Acad. Sci. U.S.A.* **90**, 8424–8428 (1993).
15. Corfas, G., Rosen, K. M., Aratake, H., Krauss, R. & Fischbach, G. D. *Neuron* **14**, 103–115 (1995).
16. Challice, C. E. & Virag, S. *Tissue & Cell* **6**, 447–462 (1974).
17. Meyer, D. & Birchmeier, C. *Nature* **378**, 394–398 (1995).
18. Lee, K.-F. et al. *Nature* **378**, 386–389 (1995).
19. Lumsden, A. & Keynes, R. *Nature* **337**, 424–428 (1989).
20. Lumsden, A., Sprawson, N. & Graham, A. *Development* **113**, 1281–1291 (1991).
21. Moody, S. A., Quigg, M. S. & Frankfurter, A. *J. comp. Neurol.* **279**, 567–580 (1989).
22. Schneider-Maunoury, S. et al. *Cell* **75**, 1199–1214 (1993).
23. Swiatek, P. J. & Gridley, T. *Genes Dev.* **7**, 2071–2084 (1993).
24. Nieto, M. A., Bradley, L. C. & Wilkinson, D. G. *Development (Suppl.)* **2**, 59–62 (1991).
25. Covell, D. A. Jr & Noden, D. M. *J. comp. Neurol.* **286**, 488–503 (1989).
26. Pöpperl, H. et al. *Cell* **81**, 1031–1042 (1995).
27. Chang, H. & Gilbert, W. *Soc. Neurosci. Abstr.* **20**, 1694 (1994).
28. Wilkinson, D. G. & Nieto, M. A. *Meth. Enzym.* **225**, 361–373 (1993).
29. Goulding, M. D., Chalepakis, G., Deutsch, U., Erselius, J. R. & Gruss, P. *EMBO J.* **10**, 1135–1147 (1991).
30. Lai, C. & Lemke, G. *Neuron* **6**, 691–704 (1991).
31. Dent, J. A., Polson, A. G. & Klymkowsky, M. W. *Development* **105**, 61–74 (1989).
32. Simon, H. & Lumsden, A. *Neuron* **11**, 209–220 (1993).

ACKNOWLEDGEMENTS. We thank D. Ortuño, C. Jackson and P. Burrola for technical support; J.-L. Goergen for help with blastocyst injections; M. Bronner-Fraser, D. Wilkinson and K. Chien for expert advice on the developing hindbrain and heart; C. Birchmeier, D. Meyer and K.-F. Lee for communicating results prior to publication; and M. Goulding and C. Kintner for reviews of the manuscript. This work was supported by grants from the NIH (G.L. and C.L.), a postdoctoral fellowship from the Swiss National Science Foundation (M.G.), and a predoctoral fellowship from the University of Milan (D.O.).

Requirement for *neuregulin* receptor *erbB2* in neural and cardiac development

Kuo-Fen Lee, Horst Henrich, Hua Chen*, Brian Bates†, Mien-Chie Hung* & Christopher Hauser

The Clayton Foundation Laboratories for Peptide Biology, The Salk Institute, 10010 North Torrey Pines Road, La Jolla, California 92037, USA

* Department of Tumor Biology, M. D. Anderson Cancer Center, University of Texas, Houston, Texas 77030, USA

† Whitehead Institute, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA

THE receptor *erbB2/neu* is a member of the epidermal growth factor receptor (EGFR or *erbB*) family that also includes *erbB3* and *erbB4*¹. Amplification of the *erbB2/neu* gene is found in many cancer types and its overexpression is correlated with a poor prognosis for breast and ovarian cancer patients². Investigation of the biology of *erbB2* led to the identification of a family of ligands termed *neuregulins* which included the *neu*-differentiation factors^{3,4}, the heregulins⁵, a ligand with acetylcholine-receptor-inducing activity⁶ and glial growth factor⁷. Several lines of evidence

suggest that heterodimerization of *erbB2* with other *erbB* receptors is required for neuregulin signalling¹. Here we investigate the developmental role of *erbB2* in mammalian development in mice carrying an *erbB2* null allele. We find that mutant embryos die before E11, probably as a result of dysfunctions associated with a lack of cardiac trabeculae. Development of cranial neural-crest-derived sensory ganglia was markedly affected. DiI retrograde tracing revealed that the development of motor nerves was also compromised. Our results demonstrate the importance of *erbB2* in neural and cardiac development.

Expression of both *erbB2* and neuregulin is detected in cell types derived from all three germ layers, including those in the developing nervous system^{7,14}. To determine the developmental role of *erbB2*, we generated mice carrying a deletion in the *erbB2* gene as described in Fig. 1a, b and used western blot analysis to establish whether the introduced deletion resulted in a null mutation (Fig. 1c). p185^{erbB2} was detected in +/+ but not in -/- embryos at day 10.5 (E10.5), and the level of p185^{erbB2} in +/- embryos was half of that in +/+ embryos (Fig. 1c). These results show that the deletion in *erbB2* leads to a null mutation.

Both *erbB2* and neuregulins are found in the neural crest (NC) cells that migrate out of the neural tube, suggesting that both molecules act via an autocrine mechanism to play an important role in the development and differentiation of NC cells. In the peripheral nervous system, trunk NC cells give rise to the neuronal cells of the dorsal root ganglion (DRG), whereas cranial sensory neurons are derived from cranial NC cells and placodal ectoderm^{15,16}. NC-derived and placode-derived neurons occupy different anatomical locations. In the trigeminal ganglion (Vg), NC-derived neurons are located in the dorsomedial (DM) region and placode-derived neurons occupy the ventrolateral (VL) region¹⁵. In the facial ganglion (VIIg), the NC-derived neurons that represent a small component of VIIg are located only in the proximal portion of the VIIg nerve root. In the glossopharyngeal

(IXn) and vagus (Xn) nerves, NC-derived superior and jugular ganglia occupy the proximal parts, while placode-derived petrosal and nodose ganglia are located in the distal region of respective nerves¹⁵.

We investigated whether a lack of *erbB2* could lead to abnormal neural development using a whole-mount immunohistochemistry method with TuJ1 antibodies raised against neuronal tubulin. As shown in Fig. 2a, these antibodies detect cranial ganglia, DRGs and motor nerves in E10.5 embryos. When *erbB2*-mutant embryos were examined, no TuJ1 immunoreactivity was detected in either the dorsal portion of the Vg or the mandibular branch of the Vg, whereas the maxillary and ophthalmic branches were normal (Fig. 2b). Haematoxylin and eosin (H/E)-stained transverse sections of mutant embryos revealed that the DM portion of the Vg was lost but not the VL portion (compare Fig. 2c and d). In contrast to control embryos, no axonal connections between the Vg and hindbrain were evident in any sections throughout the entire ganglion (Fig. 2d). In mutant embryos, proximal portions of the IXn and Xn were considerably smaller than in control embryos, suggesting that the development of the superior and jugular ganglia was severely affected, which we confirmed by H/E staining of transverse sections (data not shown). Although the VIIg was normal in size, examination of the facial motor nerve suggested that NC-derived neurons in the VIIg were also affected (see below). Flat-mounts of dissected hindbrain region showed no alteration in the number or gross structure of rhombomeres in mutant embryos (data not shown). In the trunk region, DRG sizes were similar in mutant and control embryos examined by whole-mount immunohistochemistry (Fig. 2a, b) and in H/E-stained sections (Fig. 2e, f). Considered together with the evidence for *erbB2* expression in NC cells, these results indicate that a lack of *erbB2* affects the normal development of cranial NC-derived neurons but not placode-derived neurons at E10.5.

FIG. 1 a, Targeted inactivation of the *erbB2* gene. Top, restriction map of the region of genomic DNA containing the promoter and the first exon; middle, targeting vector in which a *Bam*HI-*Xba*I 2.0-kb fragment containing the promoter and the first exon²⁹ was replaced by a PGK-neomycin cassette. This construction contains 4.0-kb and 2.0-kb homologous fragments at the 5' and 3' ends, respectively. Bottom, structure of the mutated *erbB2* locus after homologous recombination. b, Germline transmission of the targeted mutation. Mice heterozygous for the mutation were intercrossed to generate homozygous animals. Of 9 embryos collected at E10.5, one was homozygous (lane 9), one was wild type (lane 8), and the rest were heterozygotes (lanes 1-7). c, Western blot analysis of proteins isolated from E10.5 embryos. A band of M_r 185K (p185^{erbB2}) was detected in control embryos (lane 1) but not in homozygote embryos (lane 2). The heterozygote embryos expressed *erbB2* protein at one-half the level of control embryos (lane 3).

METHODS. The targeting vector was constructed according to standard procedures. Linearized targeting vector was electroporated into J1 ES cells derived from 129-strain mice as described³⁰. Heterozygote ES cells were injected into C57/B6 blastocysts to generate chimaeras which were used to establish heterozygote mice. Phenotypes were analysed in embryos derived from intercrossing heterozygote mice on a mixed genetic background of 129 and C57/B6 strains. Western blot analysis was done as described³⁰; samples (150 µg each) were run on 6% SDS-PAGE and transferred to a nitrocellulose filter, which was then incubated with primary antibody against *erbB2* (clone Ab-3; Oncogene Science) followed by horseradish peroxidase-conjugated goat anti-mouse IgG (1:1,000; Bio-Rad). The protein was visualized with enhanced chemiluminescent detection reagents (Amersham).

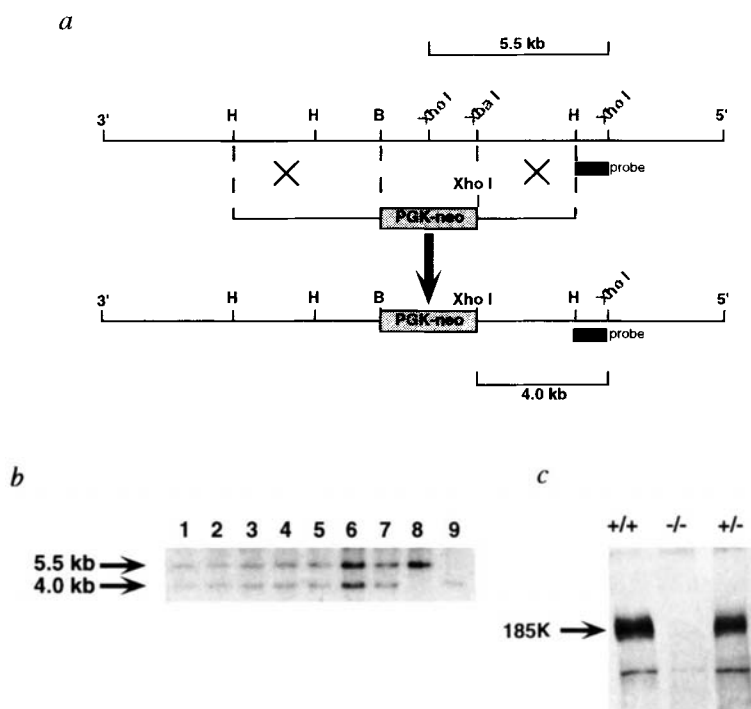
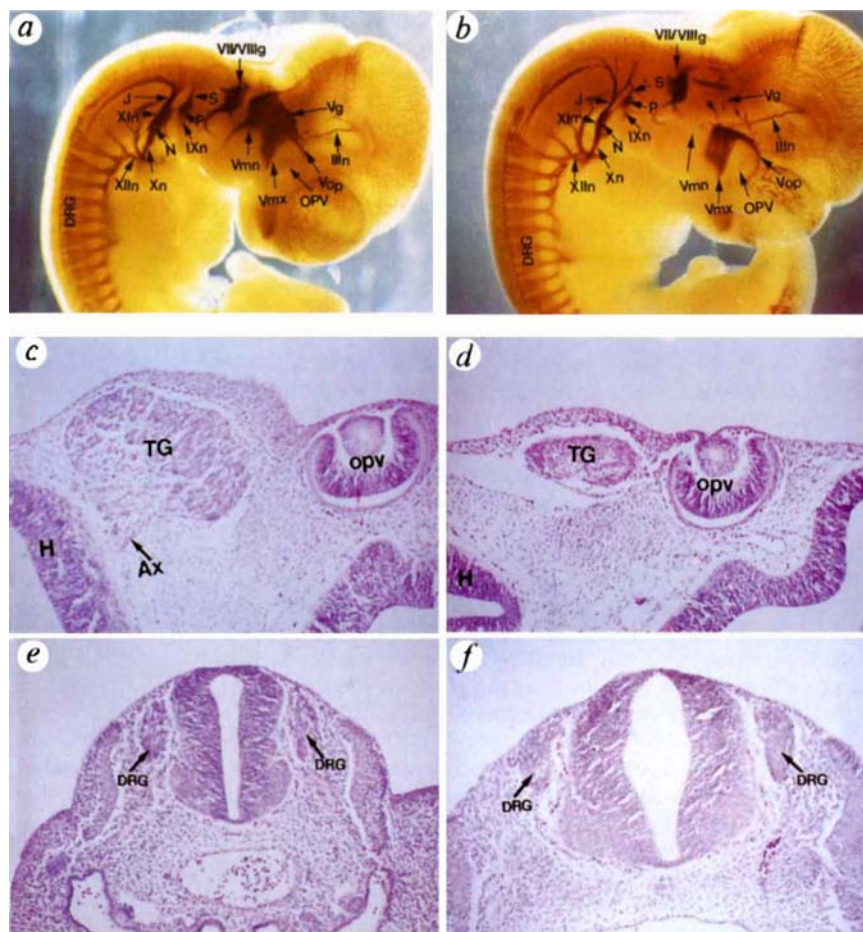


FIG. 2 Deficits in neurogenesis in *erbB2*-mutant embryos. E10.5 control (a) and mutant (b) embryos were subjected to whole-mount immunocytochemistry using monoclonal antibodies (TuJ1) against neuronal tubulin³¹. These antibodies label cranial nerves and ganglia as well as dorsal root ganglia (DRG). In mutant embryos, the proximal portion of the trigeminal ganglion was lost. The proximal parts of the glossopharyngeal nerve and vagal nerve in mutant embryos were much thinner than in control embryos, suggesting a loss of jugular and superior ganglia, respectively. Illn, oculomotor nerve; Vg, trigeminal ganglion; OPV, optic vesicle; VII/VIIIg, facial and acoustic/vestibular ganglion complex; IXn and IXg, glossopharyngeal nerve and ganglion; P, petrosal ganglion; Xn, vagal nerve; N, nodose ganglion; Xln, accessory nerve; XII, hypoglossal nerve; Vopv, Vmn, and Vmn, ophthalmic, maxillary and mandibular branches of the Vg, respectively. c–f, Histological analysis of the nervous system. Transverse sections of E10.5 control (c and e) and mutant (d and f) embryos were stained with H/E to show the morphology of Vg (c and d) and DRG (e and f). Note the loss of dorsomedial portion of Vg in mutant embryos. DRG was not reduced in size in E10.5 mutant embryos. TG, trigeminal ganglion; H, hindbrain; Ax, axon; DRG, dorsal root ganglion. METHODS. For whole-mount immunocytochemistry, embryos were collected in HEPES buffer, fixed in methanol and dimethylsulphoxide (DMSO) (4:1) at 4 °C for 4 h or overnight, and bleached with 10% hydrogen peroxide (H_2O_2) in fixative for 5 h at room temperature³². Subsequent operations were all done at room temperature. After washing with TBS (50 mM Tris and 0.85% NaCl, pH 7.5) twice for 30 min each, embryos were incubated overnight with TuJ1 antibody³¹ (1:500) in blocking solution (calf serum: DMSO, 4:1) containing 0.1% thimerosol (Sigma). After five washes with TBS 1 h each, embryos were then incubated overnight with secondary antibody (goat anti-mouse IgG-conjugated with horseradish peroxidase, Fab fragment, 1:300; Sigma). After five more washes with TBS of 1 h each, 3,3'-diaminobenzidine (DAB) (0.5 mg ml^{-1}) solution containing 0.02% H_2O_2 was added and the embryos incubated for 1.5 h; the reaction was stopped by washing



embryos with TBS and adding an equal volume of methanol. Embryos were cleared in glycerol and fixed in 4% paraformaldehyde in 0.1 M phosphate buffer. For histological analysis, embryos were dehydrated in graded alcohols and xylene, and embedded in Paraplast. Serial 7- μm transverse sections were cut and stained with Harris' haematoxylin and eosin³⁰.

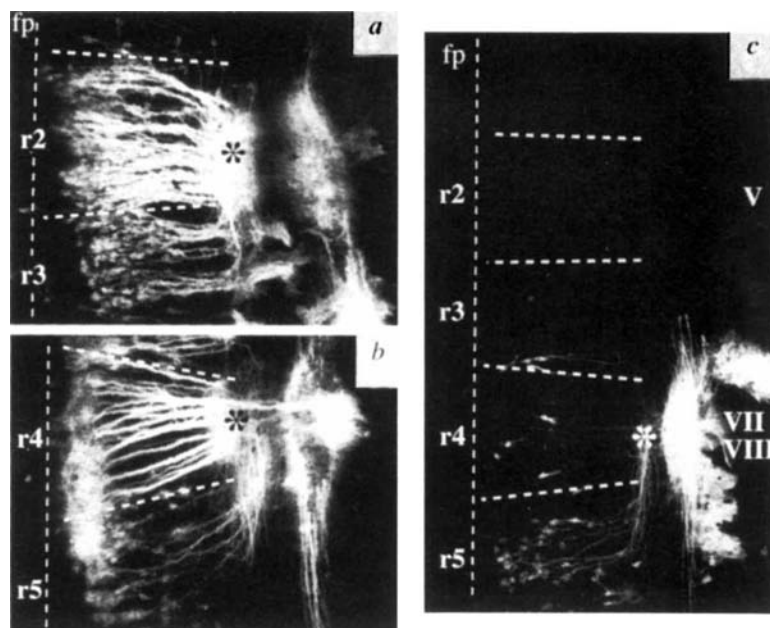


FIG. 3 Impairment of trigeminal and facial motor nerve development in *erbB2*-mutant embryos. Dil was injected into the trigeminal (a) or facial (b) ganglion of E10.5 control and mutant (c) embryos. In mutant embryos, no labelled trigeminal motor neurons were observed. Only a few labelled nerves and cell bodies were found after facial ganglion injection. r2–5, Rhombomeres 2–5; fp, floor plate; V and VII/VIII mark the location of exit/entry points for trigeminal and facial nerve/acoustic efferent, respectively; the asterisk indicates the exit point of the motor nerve. METHODS. Retrograde axonal tracing was carried out as described previously²¹. Embryos were fixed in 4% paraformaldehyde overnight and pinned down on sylgard-bottomed dishes. A solution of Dil C_{18} (1,1'-dioctadecyl-3,3,3,3'-tetramethyl indocarbocyanine; Molecular Probes D-282; 5 mg ml^{-1} in dimethyl formamide) was pressure-injected into the appropriate ganglion or nerve root. Dil images were obtained by using a Bio-Rad MRC 600 confocal microscope.

In the hindbrain, both cranial ganglia and motor nuclei are generated in a segmented fashion^{17,18} and exit points of the cranial motor nerves occupy distinct positions in relation to individual rhombomeres¹⁷. For example, motor neurons of the Vn and the VIIn are generated predominantly in r2–r3 and r4–r5 regions, respectively, whereas the axons of the V and VII motor nuclei converge towards their exit points at r2 and r4, respectively (Fig. 3a, b). Development of V motor nerves depends on the development of the Vg^{19,20}: when the Vg anlagen¹⁹ is surgically removed or in-growth of axons of Vg is prevented by a physical barrier²⁰, the V motor nerve does not develop. As NC-derived Vg was markedly affected in *erbB2*-deficient embryos, we investigated whether the motor nerves were affected. The dye DiI was injected into either the Vg or the region between the Vg and the neural tube. As indicated in Fig. 3c, no labelled sensory axons or motor neurons were seen in the r2–r3 region of mutant embryos. Similarly, when DiI was injected into the VIIg close to the exit point, the number of labelled axons and cell bodies was substantially reduced (Fig. 3c). The few labelled nerves and cell bodies in r4 and r5 may represent the efferents for the acoustic ganglion²¹. When DiI was deposited further distally and exclusively into VIIg, no labelled cell bodies were seen (data not shown). Although neuregulin mRNA was detected in most, if not all, motor nuclei¹³, the development of nerves III and XII (Fig. 2a, b) and of spinal motor nerves (data not shown) of mutant embryos was normal. It is likely that the observed deficits in the development of trigeminal and facial motor nerves are secondary to a lack of the NC-derived portion of their respective sensory ganglia^{19,20}.

Both *erbB2* and neuregulins are expressed in many non-neuronal tissues. In E9.5/10.5 embryos, for example, *erbB2* is detected in cardiac myocytes (Fig. 4a), whereas neuregulins are expressed in the adjacent endocardium¹³. These results suggest that *erbB2* and neuregulins foster development of the heart via a paracrine mechanism. During analysis of the nervous system of *erbB2*-mutant embryos, no viable embryos were found after E11. There was a complete absence of trabeculae in the mutant embryos (Fig. 4b, c). As the spongy, trabecular myocardium is largely responsible for the maintenance of blood flow during early stages of heart morphogenesis before expansion of the compact zone, the most likely cause of embryonic mortality is defective cardiac function associated with a lack of trabeculae. The endocardial cushion was normal or reduced in size in some embryos (Fig. 4a, c), but the compact zone and the outflow tract were normal (data not shown).

It is not known which developmental mechanisms are affected by a null mutation in *erbB2*. During embryonic development of the mouse trigeminal system, NC cells begin to colonize their terminal location by E9.5, where they proliferate and differentiate into neurons which extend their axons centrally and peripherally. Neuregulins have pleiotropic activities in different cell types, including differentiation^{6,12,22}, proliferation⁷ and survival^{23,24}. Thus interaction between neuregulins and *erbB2* may influence NC cell migration, NC cell proliferation, and survival of NC-derived neurons. There may be an increased incidence of cell death of trigeminal NC cells in E9.5 mutant embryos (our unpublished results). In contrast to cranial ganglia, the size of DRGs in *erbB2*-deficient embryos is normal at E10.5, so compensatory mechanisms may promote DRG neuronal survival. Alternatively, as NC cells contributing to the cranial ganglia migrate out of the neural tube earlier than those contributing to the DRG^{25,26}, DRG neurons could develop normally up to E10.5 and then degenerate. Our results cannot rule out the possibility that the development of glial cells in the DRG was affected because neuregulin induces trunk NC cells in culture to differentiate into glial cells¹². A reliable marker for glial cells at this stage of development will be required to address this issue. In the heart, the loss of trabeculation is probably due to a lack of myocyte proliferation or maturation via a paracrine mechanism.

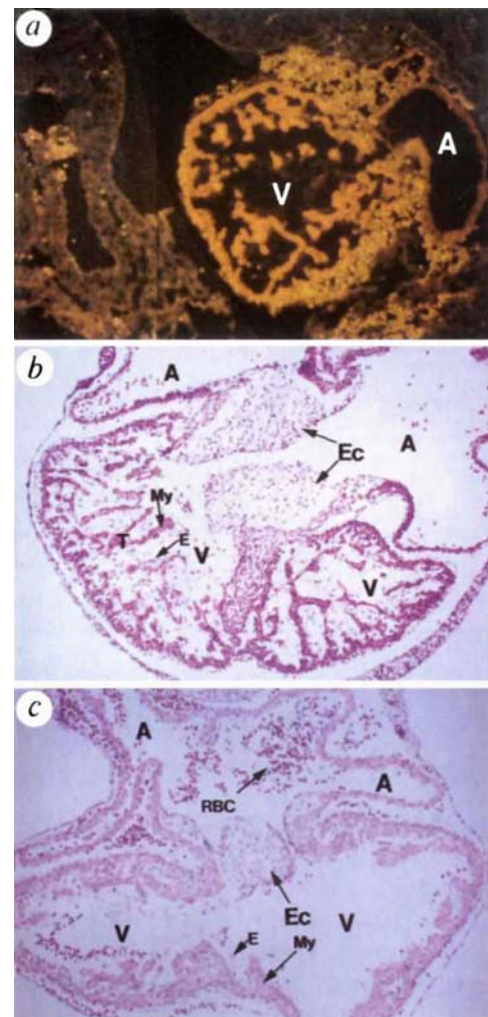


FIG. 4 Expression of *erbB2* is localized in the myocytes of developing heart and trabeculation is impaired in *erbB2*-mutant embryos. a, Sagittal section of E10.5 wild-type embryo was doubly immunostained with an anti-*erbB2* antibody and a ventricle-specific myosin-light-chain antibody. Orange signal indicates myocytes were labelled by both antibodies; yellow is background signal from red blood cells. Histological analysis of transverse sections revealed that no trabeculae were formed in E10.5 mutant embryos (c) compared to controls (b). A, atrium; V, ventricle; My, myocyte; Ec, endocardial cushion; T, trabeculae; E, endocardium; RBC, red blood cell.

METHODS. Immunocytochemistry was done as described³⁰. Embryos were fixed in 4% paraformaldehyde, cryoprotected with 30% sucrose in 0.1 M phosphate buffer, sectioned at 10 μ m, and collected onto gelatin-coated slides. Sections were washed several times with PBS and incubated for 1 h at room temperature in 10 mM phosphate buffer containing 0.5 M NaCl, 3% BSA, 0.3% Triton X-100 and 0.1% sodium azide, then incubated overnight with rabbit primary antiserum against ventricle-specific myosin light chain (1:50 dilution; kindly provided by K. Chien, UCSD) and mouse monoclonal antibody against *erbB2* (1:300). After several washes with PBS, sections were incubated for 2 h with Texas red-conjugated donkey anti-rabbit (1:300; Jackson Immunological) and fluorescein-conjugated goat anti-mouse secondary sera (Fab fragment; 1:200; Sigma), washed with PBS, and mounted in PBS:glycerol (1:1). Histological analysis of heart tissue is as described in Fig. 2 legend.

While our study was in progress, neuregulin-deficient mice and *erbB4*-deficient mice were established^{27,28}. Neuregulin-deficient mice display a cranial neural and cardiac phenotype remarkably similar to that of *erbB2*-deficient mice²⁷ and *erbB4*-deficient mice have a lack of trabeculae just like *erbB2*-deficient mice²⁸.

Together these results suggest that both erbB2 and erbB4 are not functionally redundant and are consistent with the model that, although erbB2 alone does not bind neuregulins, heterodimerization of erbB2 with other erbB receptors is required for neuregulin signal transduction¹. These mutant mice offer a system in which to explore the role of neuregulin in mammalian development. □

Received 17 August; accepted 20 October 1995.

- Carraway, K. L. III & Burden, S. J. *Curr. Opin. Neurobiol.* **5**, 1–7 (1995).
- Hynes, N. E. & Stern, D. F. *Biochem. biophys. Acta* **1198**, 165–184 (1994).
- Peles, E. et al. *Cell* **69**, 205–216 (1992).
- Wen, D. et al. *Cell* **69**, 559–572 (1992).
- Holmes, W. E. et al. *Science* **256**, 1205–1210 (1992).
- Falls, D. L., Rosen, K. M., Corfas, G., Lane, W. S. & Fischbach, G. D. *Cell* **72**, 801–815 (1993).
- Marchionni, M. A. et al. *Nature* **362**, 312–318 (1993).
- Kokai, Y., Cohen, J. A., Drebin, J. A. & Greene, M. I. *Proc. natn. Acad. Sci. U.S.A.* **84**, 8498–8501 (1987).
- Press, M. F., Cordon-Cardo, C. & Slamon, D. J. *Oncogene* **5**, 953–962 (1990).
- Orr-Urtreger, A. et al. *Proc. natn. Acad. Sci. U.S.A.* **90**, 1867–1871 (1993).
- Meyer, D. & Birchmeier, C. *Proc. natn. Acad. Sci. U.S.A.* **94**, 1064–1068 (1994).
- Shah, N. M., Marchionni, M. A., Isaacs, I., Stroobant, P. & Anderson, D. J. *Cell* **77**, 349–360 (1994).

- Corfas, G., Rosen, K. M., Aratake, H., Krauss, R. & Fischbach, G. D. *Neuron* **14**, 103–115 (1995).
- Ho, W.-H., Armanini, M. P., Nuijens, A., Phillips, H. S. & Osheroff, P. L. *J. Biol. Chem.* **270**, 14523–14532 (1995).
- D'Amico-Martel, A. & Noden, D. M. *Am. J. Anat.* **166**, 445–468 (1983).
- Le Douarin, N. M. & Smith, J. A. *Rev. Cell Biol.* **4**, 375–404 (1988).
- Lumsden, A. & Keynes, R. *Nature* **337**, 424–428 (1989).
- Lumsden, A., Sprawson, N. & Graham, A. *Development* **113**, 1281–1291 (1991).
- Moody, S. A. & Heaton, M. J. *comp. Neurol.* **213**, 327–343 (1983).
- Moody, S. A. & Heaton, M. B. *J. comp. Neurol.* **213**, 344–349 (1983).
- Simon, H. & Lumsden, A. *Neuron* **11**, 209–220 (1993).
- Vartanian, T., Corfas, G., Li, Y., Fischbach, G. D. & Stefansson, K. *Proc. natn. Acad. Sci. U.S.A.* **91**, 11626–11630 (1994).
- Pinkas-Kramarski, R. et al. *Proc. natn. Acad. Sci. U.S.A.* **91**, 9387–9391 (1994).
- Dong, Z. et al. *Neuron* **15**, 589–596 (1995).
- Weston, J. A. & Butler, S. L. *Dev. Biol.* **14**, 246–266 (1966).
- Nichols, D. H. *J. Embryol. exp. Morph.* **64**, 105–120 (1981).
- Meyer, D. & Birchmeier, C. *Nature* **378**, 394–398 (1995).
- Gassmann, M. et al. *Nature* **378**, 390–394 (1995).
- White, M. R.-A. & Hung, M.-C. *Oncogene* **7**, 677–683 (1992).
- Lee, K. F. et al. *Cell* **69**, 737–749 (1992).
- Easter, S. S. Jr, Ross, L. S. & Frankfurter, A. *J. Neurosci.* **13**, 285–299 (1993).
- Dent, J. A., Polson, A. G. & Klymkowsky, M. W. *Development* **105**, 61–74 (1989).

ACKNOWLEDGEMENTS. We thank D. Getman, P. Gruber and K. Chien for critical comments on the manuscript, B. Coyne for excellent secretarial assistance, and M. Gassmann, G. Lemke, D. Meyer and C. Birchmeier for sharing their unpublished results. This research was supported by the Fritz B. Burns Foundation, the Joseph Drown Foundation, the Robert J. & Helen C. Kleberg Foundation, the Dextra Baldwin McGonagle Foundation Inc., the Albert G. & Olive H. Schlink Foundation, the Joseph & Yetta Weisberger Fund, and grants from the NIH (K.-F.L. and M.-C.H.). H.S. was supported by a Human Frontier Science Program postdoctoral fellowship.

A huntingtin-associated protein enriched in brain with implications for pathology

Xiao-Jiang Li^{*†‡}, Shi-Hua Li^{*†‡}, Alan H. Sharp^{*†}, Frederick C. Nucifora Jr^{*†}, Gabriele Schilling^{*†}, Anthony Lanahan^{‡§}, Paul Worley^{‡§}, Solomon H. Snyder^{†‡} & Christopher A. Ross^{*†‡||}

* Laboratory of Molecular Neurobiology, † Department of Psychiatry, ‡ Department of Neuroscience, and § Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA

HUNTINGTON'S disease (HD) is an autosomal dominant neurodegenerative disorder caused by an expanding polyglutamine repeat in the IT15 or huntingtin gene¹. Although this gene is widely expressed^{2–9} and is required for normal development^{10–12}, the pathology of HD is restricted to the brain, for reasons that remain poorly understood. The huntingtin gene product is expressed at similar levels in patients and controls, and the genetics of the disorder^{13,14} suggest that the expansion of the polyglutamine repeat induces a toxic gain of function, perhaps through interactions with other cellular proteins^{15–18}. Here we report the identification of a protein (huntingtin-associated protein (HAP)-1) that binds to huntingtin. This binding is enhanced by an expanded polyglutamine repeat, the length of which is also known to correlate with the age of disease onset^{19–21}. The HAP-1 protein is enriched in the brain, suggesting a possible basis for the selective brain pathology of HD.

A complementary DNA encoding the first 230 amino acids of huntingtin, containing 44 glutamine repeats, was ligated to the GAL-4 binding domain in plasmid vector pPC97 for yeast two-hybrid screening with a rat brain cDNA library. This region was chosen because it includes the portion of the protein in which a glutamine repeat is expanded in HD¹. Alleles with 44 repeats are among the most common in the expanded range, and can cause onset of the disorder at ages ranging from the 20s to 70s^{19–21}. Screening of approximately 100,000 colonies resulted in

one positive clone. This clone (rHAP1, rat huntingtin-associated protein1 binding region) had an insert of 507 base pairs with no homology to known genes. Co-transformation of yeast with rHAP1 plus the HD constructs resulted in blue colonies that were not evident following co-transformation with either vector alone or with unrelated fusion proteins in either vector (Fig. 1a). The interaction was positive using the HD fusion protein containing either a normal repeat length (23 glutamines) or an expanded repeat (44 glutamines). The interaction of rHAP1 with huntingtin was stronger with the 44 glutamine repeats than with the 23 repeats, as determined by a semiquantitative liquid assay (Fig. 1b). To determine whether the interaction is specific to huntingtin or would be positive for another protein with a polyglutamine tract, we performed a similar experiment using rHAP1 plus an atrophin-1 construct, which contains a 21 glutamine repeat²². This showed no evidence for interaction in either filter or liquid assays.

In vitro binding assays were then performed using glutathione S-transferase (GST) fusion proteins (Fig. 1c). An HD cDNA encoding the amino-terminal region (930 amino acids with 44 repeats) was fused in frame to an expression vector pEBVHIS (Invitrogen). Transfection of this construct into HEK-293 cells resulted in a band, of relative molecular mass 131,000 (*M_r* 131K), that was recognized by a huntingtin-specific antibody⁴ (Fig. 1d). The endogenous protein (*M_r* 350K) could also be seen. When an extract from cells transfected with the N-terminal huntingtin construct was incubated with GST-HAP1 protein linked to agarose beads, both the native huntingtin and the transfected huntingtin with the expanded repeat (44 glutamines) were specifically retained on the beads (Fig. 1c–e). A similar experiment was performed using lymphoblastoid cell lines from a normal individual and HD patients with varying lengths of glutamine repeats (Fig. 1f). Cell extracts containing the huntingtin with 82 repeats yielded stronger binding compared to the huntingtin with 44 repeats, and both proteins showed increased binding to GST-HAP1 compared to the protein from the normal individual (19, 22 repeats) (Fig. 1g). Alleles of ≥ 70 repeats are in the severely expanded range and nearly always cause HD with juvenile onset.

The cDNA insert from the rHAP1 clone was used to screen rat brain cDNA libraries to obtain a full-length cDNA sequence. Two groups of cDNAs that display different sequences in the C-terminal region are represented by rHAP1-A and rHAP1-B (Fig. 2a, b). There are no significant homologies to known genes

|| To whom correspondence should be addressed.

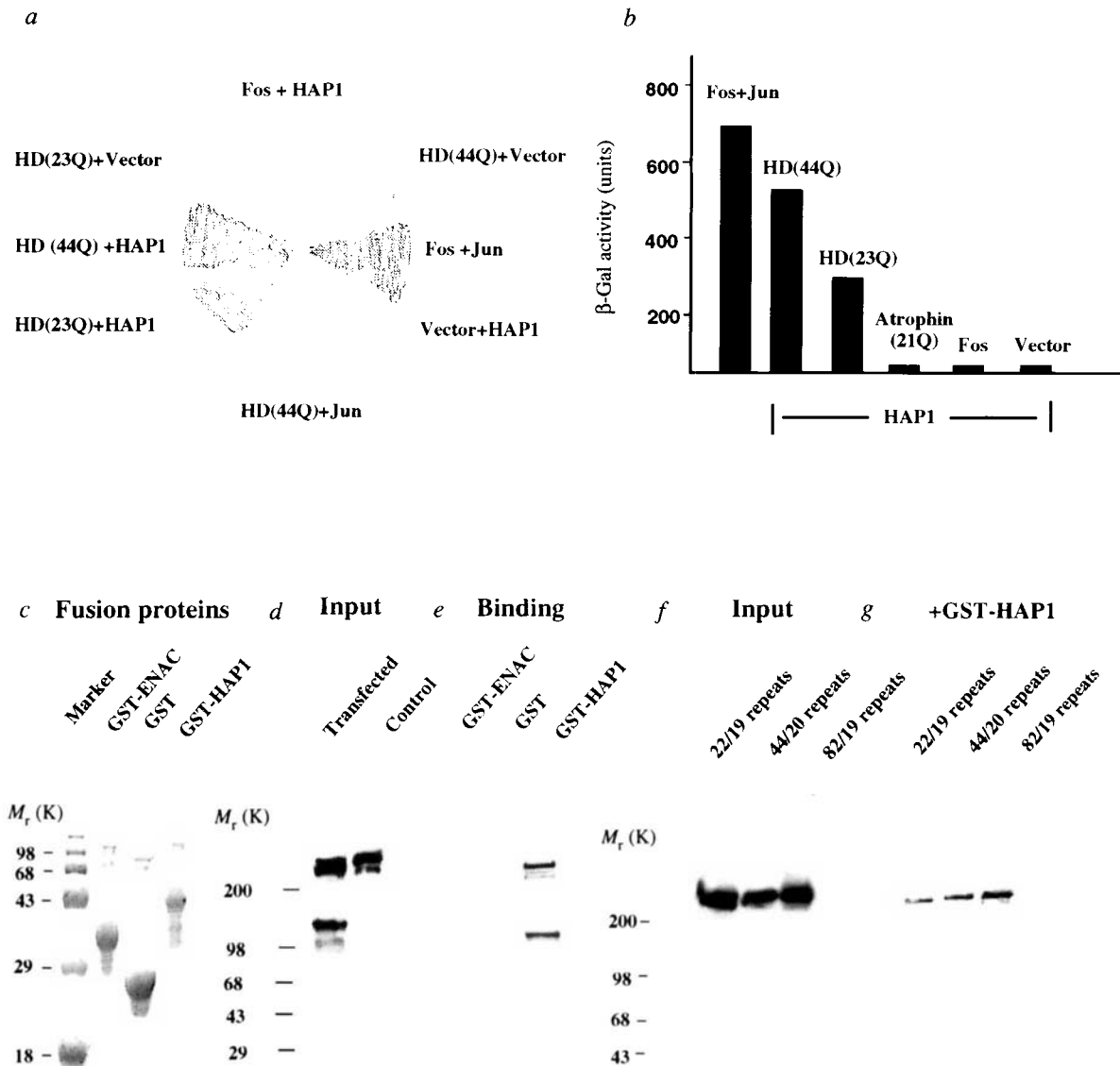


FIG. 1 Association of rat HAP1 with huntingtin. **a**, Filter assays for β -galactosidase activity in transformed yeast with proteins fused to the GAL-4 binding domain (BD) and the activation domain (AD) (BD construct + AD construct). HAP1:AD construct of rHAP1. Only coexpression of rHAP1 plus the N-terminal portions of huntingtin containing either 23 or 44 glutamine repeats (HD23Q or HD44Q) activate the β -galactosidase reporter gene (blue). Co-transformation with c-Fos and c-Jun (Fos+Jun) served as the positive control. **b**, Liquid assays of interactions of rHAP1 with various proteins in yeast cells. Atrophin (21Q) is the GAL-4 binding domain construct containing 263 amino acids of atrophin-1 with 21 glutamine repeats. **c**, Coomassie blue staining of GST-ENAC, GST and GST-HAP1 (rHAP1) fusion proteins. ENAC is the N-terminal portion (107 amino acids) of α subunit of the amiloride-sensitive sodium channel expressed predominantly in colon, kidney and tongue but not in brain^{24,25}. **d**, Western blot analysis of huntingtins from 293 cells transfected with an HD cDNA using anti-peptide antibody for the huntingtin (AP-81)⁴. **e**, Binding of GST-HAP1 to huntingtins in transfected 293 cells. **f**, Inputs of huntingtins from lymphoblasts with 22/19, 44/20 and 82/19 (two normal, or expanded and normal allele) glutamine repeats, respectively, were reacted with the GST-HAP1 fusion protein, and western blot (**g**) reveals the amount of huntingtin bound to GST-HAP1.

METHODS. The partial length of HD cDNAs were isolated from a human brain cortex cDNA library (Stratagene) using a CAG repeat oligonucleotide probe²⁶ or PCR with the first strand cDNA from a HD brain tissue. A cDNA construct containing the first 230 amino acids with 44 or 23

glutamine repeats was in-frame fused with the GAL4 DNA binding domain of plasmid pPC97 (ref. 27). An adult rat hippocampus cDNA library was constructed using plasmid pPC86 containing the GAL4 activation domain. Transformations of yeast strain PCY2 and filter assays of β -galactosidase activity were performed as described previously²⁷, and the controls were included as described²⁸. The liquid assay to quantify β -galactosidase activity is as previously described²⁹. rHAP1 cDNA (507 bp) isolated from the yeast two-hybrid screening was inserted into pGEX-2T vector (Pharmacia) and GST fusion protein expression and purification were essentially as previously described³⁰. The sources containing huntingtins were 293 cells transfected with the cDNA encoding the partial N-terminal portion (930 amino acids with 44 glutamine repeats) of huntingtin fused in frame to pEBVHIS vector (Invitrogen) and transformed lymphocytes (lymphoblast) from a normal individual and HD patients. The 293 cells from a 10-cm dish or 100-ml lymphoblast cultures were lysed in 1.5-ml binding buffer (1% Triton, 160 mM NaCl, 50 mM HEPES, pH 7.4, 2.5 mM MgCl₂, 1.5 mM CaCl₂ and 2.5 mM KCl, 1 mM PMSF, and 2 μ g ml⁻¹ each of pepstatin, leupeptin and aprotinin). Cell lysates were preincubated with agarose beads containing about 0.4–1 μ g GST protein at 4 °C for 30 min, clarified by centrifugation at 10,000 g for 10 min at 4 °C, then incubated with agarose beads containing 0.5 μ g GST-HAP1 or other fusion proteins at 4 °C for 1 h. Agarose beads were washed three times with 1 ml binding buffer. Proteins eluted from beads with SDS-sample buffer were detected in western blot analysis with huntingtin antibodies AP-81.

but display different deletions or insertions in the region corresponding to the midportion of rHAP1. These cDNAs are termed hHLP (human HAP-like protein) because they all lack the midportion of rHAP1 (between Gly 371 and Asp 421) (Fig. 2c) and were unable to generate blue colonies when introduced into yeast with HD cDNA.

FIG. 3 Expression of HAP1. *a*, Northern blot analysis of rat HAP1 mRNA, and *b*, RT-PCR analysis of expression of human HAP1 using primers spanning the binding region. St.N., subthalamic nucleus; CD, caudate; Ctx., cerebral cortex; Hip., hippocampus; Cereb., cerebellum; Genomic human genomic DNA; dH2O., no RNA input; hHAP1, hHAP1 plasmid cDNA. *c*, Immunoblot analysis of HAP1 protein. Human Ctx., human cerebral cortex.

METHODS. Total RNA (20 µg) from each tissue and the ³²P-labelled cDNA probe of rHAP1 were used in northern blot analysis, and the blot was exposed to X-ray film for 24 h. For the human RT-PCR analysis (*b*), PCR products obtained using the sense primer ATGCTCATTCTGGAGTGTGTG (MLILECV) and antisense primer AGAACCTGCAGGCATTGAGG (SMPAGS) at an annealing temperature of 62 °C for 35 cycles were separated on 1% agarose gel, transferred onto a nitrocellulose membrane, and hybridized with the ³²P-labelled cDNA probe of hHAP1 under high-stringency conditions. One of five independent RT-PCR assays with commercial human cDNA (Clontech) or our synthesized cDNA is shown in *b*. This experiment is representative, but other experiments did yield a detectable band in other brain regions, such as hippocampus. For the western blots (*c*), tissues were homogenized in PBS buffer with protease inhibitors as described in Fig. 2 and about 75 µg proteins were used in western blot except for transfected cells (1 µg). HEK-293 cells were transfected with a pCIS-2 expression vector (Genentech) containing a rat HAP1 cDNA construct using the Ca²⁺-phosphate precipitation method. The N-terminal sequence of rHAP1-A was inserted into rHAP1-B using the NcoI restriction site to generate the full-length protein of rHAP1-B. The GST-HAP1 fusion protein was used as an antigen for production of antiserum from rabbits (HRP Inc.). The antiserum (1.5 ml) was incubated with 20 µg GST-HAP1 antigen transferred onto the nitrocellulose. The nitrocellulose was washed with PBS, and antibodies were eluted with 1 ml 100 mM glycine (pH 2.8). The elution fluid was dialysed against PBS overnight. Purified antiserum (1:2,000) or preabsorbed antiserum in PBS containing 3% BSA was used for western blot analysis with an enhanced chemiluminescence kit (Amersham).

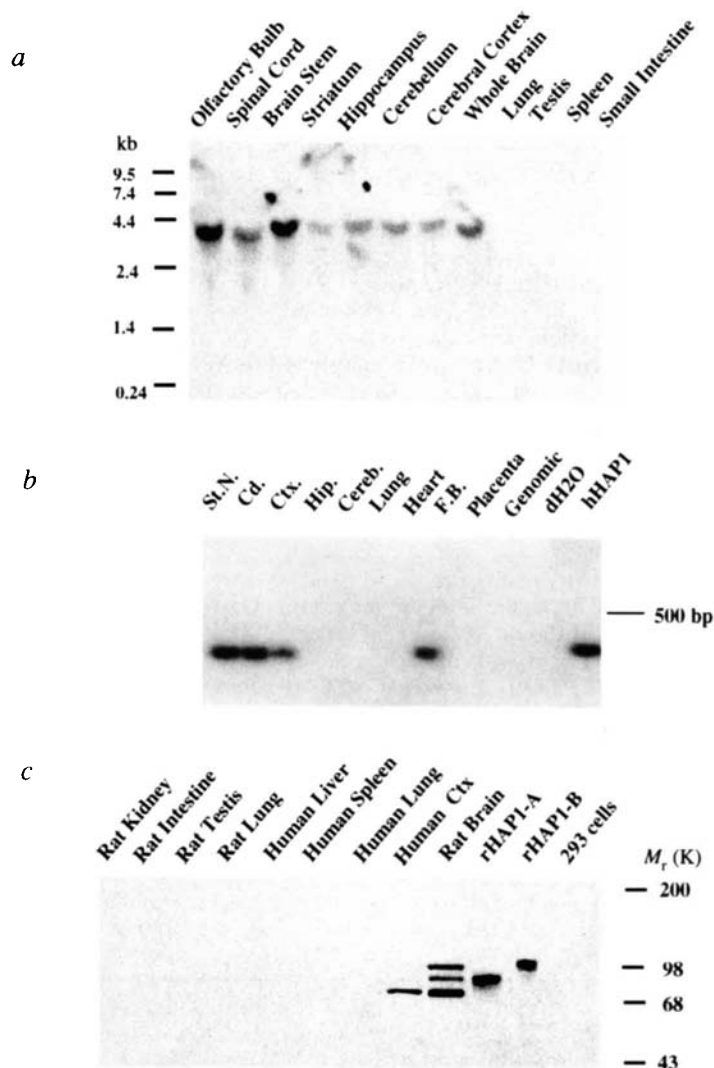
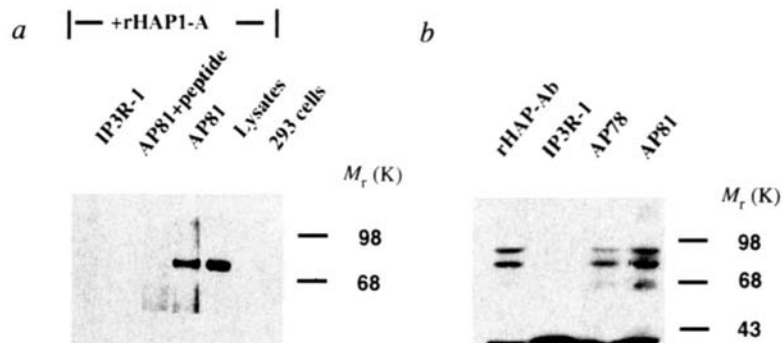


FIG. 4 Co-immunoprecipitation of *a*, transfected rHAP1-A and endogenous human huntingtins in HEK 293 cells, and *b*, endogenous rat HAP1 and huntingtins. 293 cells, untransfected cells; Lysates, cell extracts from transfected 293 cells with rHAP1-A cDNA; AP81, with purified antibody (AP-81) for huntingtin; IP3R-1, with antibody for InsP₃ receptor type 1; AP81 + peptide, with AP81 preabsorbed with the peptide immunogen; HAP1-Ab, rat HAP1 antibodies.

METHODS. Transfected cells with rHAP1-A cDNA from a 10-cm plate or whole rat brain (0.15 w/w) were lysed in 1.5 ml binding buffer as for GST fusion protein binding experiments, and 750 µl lysate was incubated with 100 µl antibodies linked beads (50% slurry) at 4 °C for 1 h. Affinity-purified antibodies (AP-78 or AP81, 20 µg) were coupled to 1 ml protein-A agarose beads using 100 mM dimethyl pimelimidate. AP81 linked beads pre-absorbed with peptide immunogen (20 µg ml⁻¹) overnight or antibodies linked beads for InsP₃ receptor type 3 were used as the controls. Proteins were eluted from beads with 100 mM glycine (pH 2.8) following washes with the



binding buffer, and eluted proteins were detected on western blots with purified rHAP1 antibodies (1:2,000). GenBank accession numbers for the cDNA sequences of rHAP1-A, rHAP1-B, hHAP1 and hHLP1 are U38370–U38373.

structs (44 or 23 repeats) produced blue colonies with similar intensity to those with rHAP1 cDNA (data not shown), indicating that the mid-portion of HAP1 is crucial for interactions with the HD protein. Southern analysis using human genomic DNA is consistent with the existence of at least two separate genes related to rHAP1 (data not shown).

Northern blot analysis of the expression of the rHAP1 (Fig. 3a) shows transcripts of approximately 4.0 kb in rat brain, with the suggestion of two partly resolved bands. Expression is high in many brain regions. By contrast there is no detectable signal in several peripheral tissues (Fig. 3a). To confirm the expression of hHAP1 in human brain, we performed reverse transcription (RT)-PCR analysis with primers derived from hHAP1 (Fig. 3b). We obtained PCR products, with the expected size (320 bp), that could be detected by hybridization with a hHAP1 cDNA probe. This form is detected in subthalamic nucleus, caudate, cerebral cortex and fetal brain.

Affinity-purified antibodies raised against GST-HAP1 were used to confirm expression of the rHAP1 protein. A construct containing the full-length protein of rHAP1-A or rHAP1-B was transfected into HEK-293 cells. Western blots of rat whole brain contain three immunoreactive bands (Fig. 3c). The middle band is the same molecular mass as the transfected rHAP1-A protein (M_r 75K) in 293 cells. *In vitro* transcription and translation of rHAP1-A cDNA also yields a product of M_r 75K (data not shown). The upper band (M_r 85K) corresponds to the transfected rHAP1-B; the lowest, M_r 68–70K, is the size predicted for rHAP1-A or rHAP1-B cDNA, and is present in both rat and human brain. By contrast, several peripheral tissues from rat and human showed no immunoreactivity (Fig. 3c). The immunoreactivity of all three bands could be eliminated by pre-absorption of antibodies with the antigen (data not shown). The three bands in the rat brain might reflect alternative splicing, which is in accord with the existence of two distinct cDNA sequences for rHAP1. The variety of HAP cDNA homologues obtained by PCR suggests the existence of multiple human HAP1 or HAP-like proteins (HLP). Antibodies for rHAP1 might be unable to distinguish them because of their similar mobilities in the gel or different expression levels. Post-translational modification may also play a role and fits with the molecular mass of the upper two bands being greater than the predicted molecular mass.

To confirm that the full-length rHAP1 interacts with huntingtin *in vivo*, we conducted co-immunoprecipitation experiments.

The rHAP1-A protein transfected in 293 cells could be precipitated specifically with affinity-purified antibody to huntingtin, AP81, and this precipitation could be blocked by the peptide antigen (Fig. 4a). Using a different peptide antibody for huntingtin (AP78), we obtained similar results (data not shown). Using whole rat brain, we also observed co-precipitation of both rat HAP1 and huntingtins with AP78 or AP81 (Fig. 4b). We have not yet clearly shown co-immunoprecipitation from human brain tissue, which may reflect the complex mixture of hHAP1 and HLPs and their possibly different tissue localization. Another factor may be the difficulty of using human post-mortem brain tissues.

The normal function of huntingtin is unknown, although it has been proposed to play a role in microtubule-mediated transport or vesicle function^{7,9}. The normal function of HAP1 is also unknown. However, the selective expression of HAP1 in brain, in contrast to the ubiquitous distribution of HD gene expression, suggests that HAP1 may contribute to the brain-specific pathology of HD. The pathology of HD involves most or all of the brain in advanced cases, although some regions show striking early selective vulnerability. Although rHAP is highly expressed in many regions of rat brain, the message for hHAP1 can be detected most consistently by RT-PCR in the human brain regions (caudate and cortex) that are most affected in HD. Thus the levels of expression of human HAP1 in specific cell populations may influence their vulnerability. In addition, intrinsic vulnerability of different population of neurons may be a factor. For instance, medium spiny neurons of the striatum appear to be especially sensitive to metabolic stress²³.

The yeast two-hybrid assays and the *in vitro* binding data show that the association between HAP1 and huntingtin is enhanced by increasing lengths of the glutamine repeat. Binding may be influenced by the flanking amino acids, as there is no binding to the atrophin-1 protein which contains essentially the same number of glutamines (21) as the HD construct (23). The affinity of the interaction may be increased by an altered conformation of the huntingtin owing to the expanded glutamine repeat. The enhanced association of HAP1 with a huntingtin containing an expanded polyglutamine repeat could contribute to the pathophysiology of HD, consistent with the toxic gain-of-function hypothesis.

Note added in proof: The selective binding of the monoclonal antibody IC2 to expanded glutamine repeats, as reported by J.-L. Mandel's group³¹, is consistent with the conformational alteration proposed here. □

Received 19 April; accepted 9 October 1995.

- Huntington's Disease Collaborative Research Group *Cell* **72**, 971–983 (1993).
- Li, S.-H. *et al. Neuron* **11**, 985–993 (1993).
- Strong, T. V. *et al. Nature Genet.* **5**, 259–265 (1993).
- Sharp, A. H. *et al. Neuron* **14**, 1065–1074 (1995).
- Schilling, G. *et al. Hum. molec. Genet.* **4**, 1365–1371 (1995).
- Trottier, Y. *et al. Nature Genet.* **10**, 104–110 (1995).
- DiFiglia, M. *et al. Neuron* **14**, 1075–1081 (1995).
- Persichetti, F. *et al. Molec. Med.* **1**, 374–383 (1995).
- Gutekunst, C.-A. *et al. Proc. natn. Acad. Sci. U.S.A.* **92**, 8710–8714 (1995).
- Nasir, J. *et al. Cell* **81**, 811–823 (1995).
- Duyao, M. P. *et al. Science* **269**, 407–410 (1995).
- Zeitlin, S., Liu, J.-P., Chapman, D. L., Papaioannou, V. E. & Efstathiou, A. *Nature Genet.* **11**, 155–163 (1995).
- Wexler, N. S. *et al. Nature* **326**, 194–197 (1987).
- Ambrose, C. M. *et al. Somat. Cell molec. Genet.* **20**, 27–38 (1995).
- Gusella, J. F., MacDonald, M. E., Ambrose, C. M. & Duyao, M. P. *Archs Neurol.* **50**, 1157–1163 (1993).
- Ross, C. A., McInnis, M. G., Margolis, R. L., Antonarakis, S. E. & Li, S.-H. *Trends Neurosci.* **16**, 254–260 (1993).
- Albin, R. L. & Tagle, D. A. *Trends Neurosci.* **18**, 11–14 (1995).
- Ross, C. A. *Neuron* **15**, 493–496 (1995).

- Duyao, M. *et al. Nature Genet.* **4**, 387–392 (1993).
- Andrew, S. E. *et al. Nature Genet.* **4**, 398–403 (1993).
- Stine, O. C. *et al. Hum. molec. Genet.* **2**, 1547–1549 (1993).
- Margolis, R. L., Li, S.-H. & Ross, C. A. *Molec. Brain Res.* (in the press).
- Beal, M. F. *Ann. Neurol.* **38**, 357–366 (1995).
- Canessa, C. M., Horisberger, J. D. & Rossier, B. C. *Nature* **361**, 467–470 (1993).
- Li, X. J., Blackshaw, S. & Snyder, S. H. *Proc. natn. Acad. Sci. U.S.A.* **91**, 1814–1818 (1994).
- Li, S.-H., McInnis, M. G., Margolis, R. L., Antonarakis, S. E. & Ross, C. A. *Genomics* **16**, 572–579 (1993).
- Chevray, P. M. & Nathans, D. *Proc. natn. Acad. Sci. U.S.A.* **89**, 5789–5793 (1992).
- Fields, S. & Sternglanz, R. *Trends Genet.* **10**, 286–292 (1994).
- Guarente, L. *Meth. Enzym.* **101**, 181–191 (1983).
- Guan, K. & Dixon, J. E. *Ann. Biochem.* **192**, 262–267 (1991).
- Trottier, Y. *et al. Nature* **378**, 403–406 (1995).

ACKNOWLEDGEMENTS. X.-J.L. and S.-H.L. contributed equally to this work. We thank C. Stine for the lymphoblastoid cell lines from the HD patients; R. Margolis for the human atrophin-1 cDNA; T. Dawson and M. Forte for discussion; S. Fields for discussion of the yeast two-hybrid system; P. Chevray and W. Yee for assistance with the two-hybrid system; P. McHugh for guidance and support; and the Genetics Core Facility, especially R. Ashworth, for DNA sequencing and analysis. This research was supported by the NIH, a NARSAD established investigator award, a Stanley Foundation grant, and bequests from the estate of Nora Mae Gee and the DeVelbiss fund to C.A.R., and grants to S.H.S. from the W. M. Keck Foundation and the Stanley Foundation Research Awards Program. We thank the HDF for organizing helpful symposia.

Polyglutamine expansion as a pathological epitope in Huntington's disease and four dominant cerebellar ataxias

Yvon Trottier*, Yves Lutz*, Giovanni Stevanin†, Georges Imbert*, Didier Devys*, Géraldine Cancel†, Frédéric Saudou*, Chantal Weber*, Gilles David†, Laszlo Tora*, Yves Agid†, Alexis Brice† & Jean-Louis Mandel*†

* Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), CNRS, INSERM, ULP, B.P. 163, 67404 Illkirch Cédex, France
† INSERM U.289, Hôpital de la Salpêtrière, 75651 Paris Cedex 13, France

‡ To whom correspondence should be addressed

A POLYGLUTAMINE expansion (encoded by a CAG repeat) in specific proteins causes neurodegeneration in Huntington's disease (HD) and four other disorders¹⁻⁶, by an unknown mechanism thought to involve gain of function or toxicity of the mutated protein^{7,8}. The pathological threshold is 37–40 glutamines in three of these diseases, whereas the corresponding normal proteins contain polymorphic repeats of up to about 35 glutamines¹⁻³. The age of onset of clinical manifestations is inversely correlated to the length of the polyglutamine expansion. Here we report the characterization of a monoclonal antibody that selectively recognizes polyglutamine expansion in the proteins implicated in HD and in spinocerebellar ataxia (SCA) 1 and 3. The intensity of signal depends on the length of the polyglutamine expansion, and the antibody also detects specific pathological proteins expected to contain such expansion, in SCA2 and in autosomal dominant cerebellar ataxia with retinal degeneration, whose genes have not yet been identified⁹⁻¹³.

The target proteins defective in the five known glutamine-repeat disorders—HD, spino-bulbar muscular atrophy (SBMA), SCA1, dentatorubral-pallidoluysian atrophy (DRPLA), and Machado-Joseph disease (SCA3)—are dissimilar except for the polyglutamine stretches, and probably have different functions. The function of only one of them is known, the androgen receptor implicated in SBMA¹. A polyglutamine stretch is present in other proteins, including some transcription factors¹⁴, without documented expansion. The most common allelic form of the TATA-binding protein (TBP), a basic transcription factor, has a stretch of 38 Glns^{15,16}, and monoclonal antibody (mAb) 1C2 raised against TBP recognizes peptides overlapping this polyglutamine stretch¹⁷ (Y.L., unpublished results).

To determine whether this mAb can detect other proteins containing a polyglutamine stretch, we tested, on western blot, lymphoblastoid cell lines (LCL) from normal and HD individuals, presenting various lengths of polyglutamine in the HD protein (HDP). When probed with anti-HDP mAbs¹⁸, the mutated HDP containing the expanded polyglutamine and the normal one can be discriminated on western blot (Fig. 1a). When the same blot was probed with 1C2, only the mutated HDPs were detected,

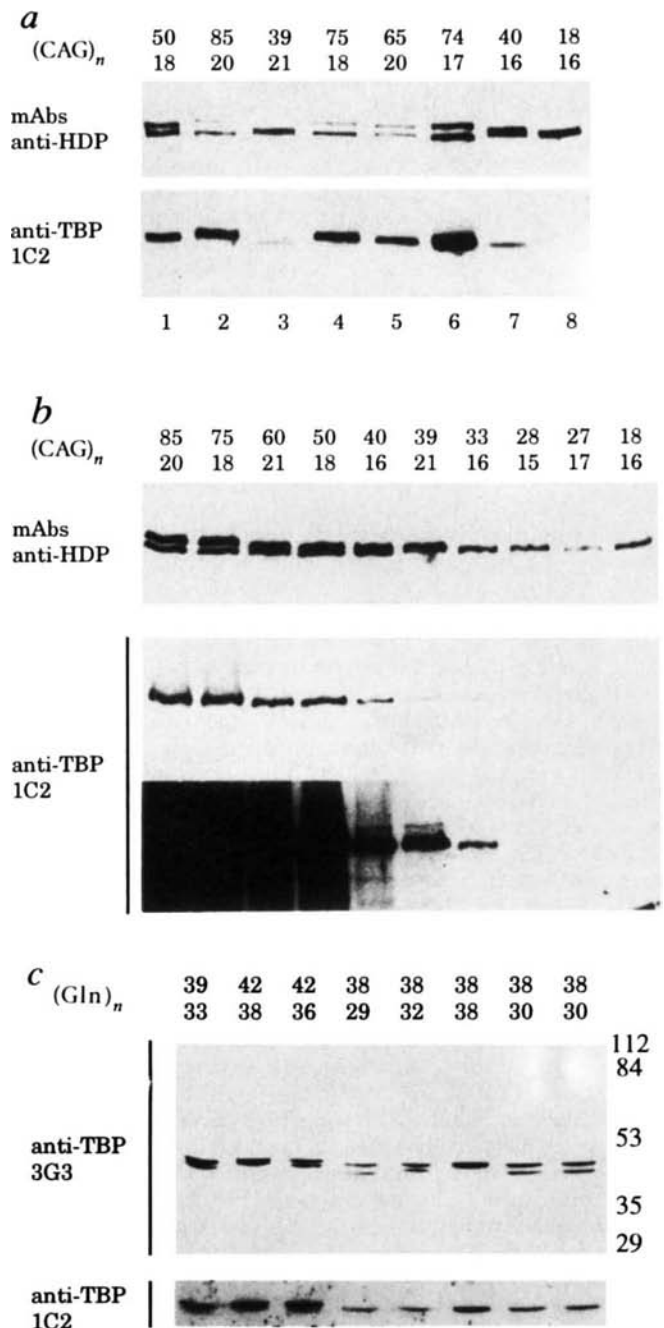


FIG. 1 Selective detection of HDP with expanded polyglutamine and of allelic forms of TBP. **a**, Western blot analysis of allelic HDPs successively immunoprobed with anti-HDP mAbs and the mAb1C2. The mutated and normal HDP are detected as a doublet with anti-HDP mAbs when the difference in the number of CAG repeats between alleles is ≥ 24 repeats (lanes 1, 2, 4–7). Only the mutated HDPs are detected by the mAb1C2. **b**, LCL extracts from HD and normal individuals are ordered by decreasing length of polyglutamine with respect to the larger allele. Using specific anti-HDP mAbs, allelic HDPs show equivalent signal intensities. With

the mAb1C2, the intensity of signal of the mutated HDPs increases with the length of the polyglutamine stretch, from 39 to 85 Glns. A gross overexposure of the blot shows a detectable signal of HDPs with 28–33 Glns (bottom). **c**, Western blot analysis of allelic TBPs (with 29–42 Glns) immunoprobed successively with the anti-TBP mAb3G3, which recognizes the N-terminal 16 amino acids of TBP¹⁷, and with the mAb1C2. Allelic TBPs are detected by 3G3 with equivalent signal, whereas 1C2 detects TBP with ≥ 33 Glns, with a stronger intensity of signal for larger polyglutamines (38–42 Glns).

METHODS. Whole-cell extracts from LCL (50 μ g of proteins) were run on SDS-PAGE gels containing either 4% polyacrylamide (with prolonged migration) for HDP analysis¹⁸ or 12% for TBP. Western blotting, immunoprobings, detection and stripping were performed as described previously¹⁸. The pool of anti-HDP mAbs was obtained by mixing supernatants of 4C8, 2E8, 2C1, each diluted at 1:100 (ref. 18). The 3G3 was ascites fluid (diluted at 1:10,000)¹⁷. After stripping out these antibodies, the blots were then probed with mAb1C2 (ascites fluid dilution 1:2,000). The estimated number of CAG repeat ((CAG)_n for HD alleles) or glutamine repeat ((Gln)_n, of the TBP alleles) was determined by PCR analysis^{15,27}, or was communicated by D. Rubinsztein for the larger normal HD alleles (Fig. 1b) in cell lines from CEPH families.

with signal intensities varying with the length of the polyglutamines (Fig. 1a, bottom). The ability of the mAb1C2 to detect mutated HDP depends on the length of the polyglutamine stretch (Fig. 1b), as proteins containing a long stretch (60–85 glutamines) gave a much higher signal intensity than those with 39 and 40 Glns (representing the smaller pathological alleles). Moreover, HDP having polyglutamine stretches at the upper normal range (≥ 28) are barely detectable after a gross overexposure (Fig. 1b, bottom). A semi-quantitative comparison of the intensity of signal using HDPs having 39, 60 and 85 Glns was performed with serial dilutions of LCL extracts. The intensity of signal of HDP with 85 Glns is 2–4 times higher than that with 60 Glns, and the latter is 10–20 times higher than HDP having 39 Glns (data not shown).

As the polyglutamine stretch in TBP is also highly polymorphic in the normal population, ranging from 25 to 42 Glns^{15,16}, we tested the ability of 1C2 to detect different TBP alleles. When probed with another anti-TBP mAb¹⁷, allelic forms of TBP containing 29–42 Glns were discriminated on western blot and gave equivalent signal intensities (Fig. 1c). However, mAb1C2 detected TBP with a long polyglutamine stretch (38–42 Glns) more strongly than those with 32 or 33 Glns, and those with only 29–30 Glns were barely detectable. Thus, in both HDP and TBP, the epitope recognized by mAb1C2 appears to be pure polyglutamine, and seems to be dependent on its length.

We therefore tested LCL extracts from SCA1 and SCA3 patients on western blot. When probed with mAb1C2, the SCA1 protein (ataxin-1) was specifically detected in SCA1 LCLs having 55 or more Glns, but was undetectable in the SCA3 LCLs (Fig. 2). Inversely, at least three SCA3 proteins (a major band at M_r 68K and two minor bands at 74K and 87K) were detected in SCA3 LCLs, but were absent in SCA1 LCLs. In brain cortex from a SCA3 patient, only the major isoform (68) was detected, and no signal was found in two normal brain samples (Fig. 2).

Mutated ataxin-1 was not detected in SCA1 LCLs with smaller expansion (42–49 Glns; Fig. 2, lanes 4–6). When compared to the signal intensity obtained for HDP having a similar length of glutamine stretch, this suggests that either the mAb1C2 detects the polyglutamine in ataxin-1 less well than in HDP or, more likely, that the expression level of ataxin-1 in LCL is lower than that of HDP.

The adult onset and the anticipation phenomenon (and its parental sex bias) characteristic of the five Gln-repeats diseases are also seen in other neurological disorders, implying possible involvement of polyglutamine expansion in these diseases. We therefore analysed patients having autosomal dominant cerebellar ataxias (ADCA) and familial spastic paraplegia (AD-FSP) that present evidence for anticipation, but for which the gene has not been identified yet. Six LCLs from ADCA and one from FSP patients were tested blind with control LCLs from HD, SCA1 and SCA3 patients. A specific protein of either 130K or 150K was detected by mAb1C2 in four patients, but the other three showed no specific signal (data not shown).

The 130K protein was found in two unrelated patients having ADCA with retinal degeneration (ADCA type II/SCA7) (Fig. 3b). The same protein was thereafter detected in patients from two other ADCA II families (data not shown). Three of these families show a strong anticipation (mean of 17 ± 14 years between parent-offspring pairs) with parental sex bias, and their linkage analysis mapped the defective gene on chromosome 3p12-p21.1 (ref. 9), in agreement with other linkage data on ADCA type II¹⁰.

The 150K protein was detected in two unrelated patients having an uncharacterized ADCA, although mutations at the SCA1, SCA3 and DRPLA loci have been excluded by direct analysis of CAG repeats. We further analysed their families and two other ADCA patients, one of them having the defective gene mapping to the SCA2 locus (12q23-q24)^{11,31} (Fig. 3a). All patients, including the SCA2 patient (lane 9), showed the 150K protein, which was not detected in a normal relative (lane 4).

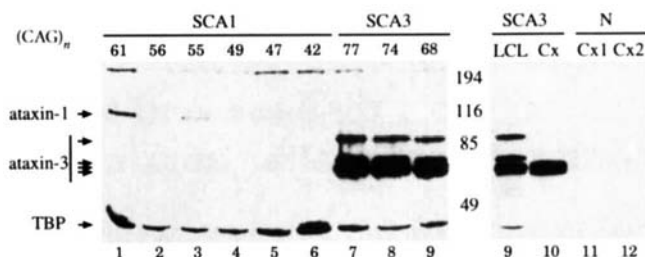


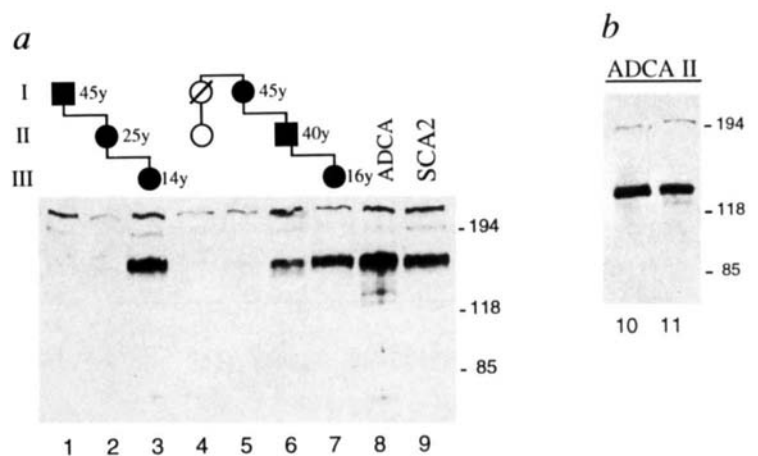
FIG. 2 Detection of the SCA1 and SCA3 pathological proteins. The mAb1C2 specifically detects ataxin-1 at ~100K (as previously described²⁸) in SCA1 LCLs with a large CAG expansion (≥ 55 CAG) (lanes 1–3), and three pathological proteins at 87K, 74K and 68K (a 64K band is also detected, but may correspond to a degradation product) in SCA3 LCLs (lanes 7–9). Ataxin-1 is not detected in this experiment when the expanded polyglutamine has ≤ 49 Glns (lanes 4–6). The ~100K protein and the SCA3 specific proteins were not detected in 9 other normal control cell lines (not shown). The TBP and a protein of M_r ~230K are detected in all LCLs (controls and patients) analysed, with a rather constant signal intensity. In cortex (Cx) or a SCA3 patient with a pathological allele with 74 CAG repeats (lane 10), only the 68K isoform (the major form in LCL) is detected, but it is absent in the two normal cortex (lanes 11 and 12). The previously cloned SCA3 cDNA⁶ was predicted to encode a 39.5K protein containing a polyglutamine stretch. The other cloned cDNA isoforms did not encode the polyglutamine⁶. The three SCA3 isoforms detected by 1C2 must contain a polyglutamine stretch, and may result from alternative splicing and/or post-translational modifications. Sequence analysis of a reverse transcription (RT)-PCR-derived cDNA clone and of a cloned PCR fragment amplified from genomic DNA (from a different individual) revealed a change of one base pair (not shown) with respect to the published sequence⁶ that modifies the predicted termination codon in a tyrosine codon, resulting in a C-terminal extension of the putative protein of 16 amino acids. Thus the M_r of the predicted SCA3 protein with ~70 Glns would be ~47K, and it might correspond to one of the isoforms detected, as proteins containing a long polyglutamine stretch migrate with apparent higher M_r than expected in SDS-PAGE gel²⁸. METHODS. Western blots (from SDS-PAGE (8%) gels) were immunoprobed with 1C2 (dilution 1:2,000). The number of CAG repeats of the pathological allele at the SCA1 and SCA3 loci (given above each lane) was estimated by PCR analysis^{3,6}.

SCA2 belongs to type I cerebellar ataxia, a group of autosomal dominant diseases (including SCA1–5) that cannot be distinguished clinically. In the two families analysed, the intensity of the signal increased in successive generations, and appears inversely correlated with the age of onset of the disease in affected individuals. By analogy with data obtained with mutated HDPs (Fig. 1), these results suggest that the polyglutamine stretch contained in the 150K protein increases in size through parental transmission, giving a stronger signal, and accounts for the anticipation phenomenon observed in these families and in previous studies of SCA2 (refs 12, 13, 19).

Our results imply that polyglutamine expansions in the 150K and the 130K proteins are responsible for the SCA2 and ADCA II phenotypes, respectively. The 1C2 antibody should facilitate the cloning of the corresponding genes by allowing the purification of the relevant proteins, the screening of cDNA expression libraries, or by improving linkage analysis as small families can be now unequivocally identified as SCA2 or ADCA II. It should also find diagnostic applications in patients with uncharacterized spinocerebellar ataxias. The 1C2 antibody will help test the hypothesis of a poly (CAG : glutamine) expansion in other neurodegenerative diseases with proven or suggested anticipation, such as other type I ADCAs (SCA4 (ref. 20) and SCA5 (ref. 21)) and AD-FSP^{22–24}. Anticipation may even occur in some families with bipolar affective illness or schizophrenia²⁵. We found no evidence for expanded polyglutamine stretch in the LCLs we analysed from patients with the AD-FSP form linked to chromosome 2 (ref. 23) or with AD-FSP form not linked to

FIG. 3 Evidence of polyglutamine expansion in SCA2 and ADCA II patients. *a*, A 150K protein is specifically detected by 1C2 in LCLs from a SCA2 patient (lane 9) and in seven patients with an uncharacterized ADCA (lanes 1–3 and 5–8), and is not detected in an unaffected relative (lane 4). Patients in lanes 3 and 7 are those for which the 150K protein was initially detected in a blind test (see text) and are presented here with their relatives. In these families, the intensity of signal of the 150K protein increases from one generation to the other, and in association with the decreasing age (in years, y) of onset of affected individuals. Retrospective analysis of linkage data for families of patients showing the 150K protein gave a maximal lod score of 4.22 ($\theta=0.00$) with marker D12S1329, and thus establish linkage to the SCA2 locus. *b*, A 130K protein is specifically detected by 1C2 in LCLs of two unrelated ADCA II patients, initially in a blind test (see text).

METHODS. Western blots (from SDS-PAGE (8%) gels) were immunoprobed with mAb1C2 (dilution 1:2,000).



chromosome 2 (data not shown). However, as false negative findings may be obtained if pathological protein has expansion too small to be detected with high sensitivity, or if it is absent in the cells analysed, further analysis of early-onset cases or pathological tissues is warranted.

The 1C2 antibody was used to characterize the subcellular localization of the proteins involved in SCA2, SCA3 and ADCA II. The pathological proteins in SCA3 (not reported previously) and the 150K protein specific to SCA2 as well as the mutated HDP, were found in the cytoplasmic fraction, while the pathological protein of 130K detected in ADCA II localized to the nucleoplasm (Fig. 4), further indicating that polyglutamine-containing proteins involved in these neurodegenerative diseases have very little in common in terms of their function.

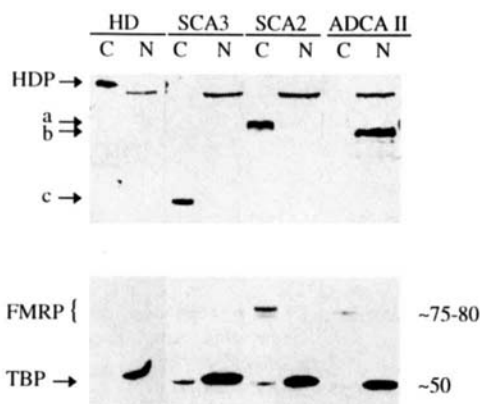


FIG. 4 Subcellular localization of pathological proteins corresponding to HD, SCA2, SCA3 and ADCA II. On western blot, nuclear (N) and cytoplasmic (C) fraction from LCLs are tested with anti-TBP 3G3 and anti-FMRP 1C3 (ref. 29). TBP is mostly in the nuclear extract, whereas most, but not all, of the FMRP isoforms are found in the cytoplasmic fraction, as already observed²⁹. When probed with mAb1C2, HDP, SCA3 proteins and the 150K protein specific to SCA2 are predominantly in the cytoplasmic fraction, whereas the 130K protein specific to ADCA II and the 230K protein which is detected in all LCLs, are found in the nuclear extract. a, b, c indicate the position of the SCA2, ADCA II and SCA3 proteins, respectively.

METHODS. Fractionation of LCLs was performed as described³⁰, generating a nuclear pellet and a postnuclear supernatant. Proteins (50 µg) of each fraction were run on SDS-PAGE (7%) gels. Western blots were probed with anti-TBP mAb3G3 (dilution 1:10,000) and anti-FMRP mAb1C3 (ascites fluid dilution 1:2,000) together. After stripping out these antibodies, blots were probed with mAb1C2.

The 1C2 antibody might be used to detect specifically the pathological proteins in brain samples from patients, or in transgenic animal models of neurodegenerative disorders, and to look for possible local accumulation of these proteins (or their breakdown products containing the polyglutamine stretches predicted by models proposing that polyglutamine stretches promote oligomerization²⁶). But the most interesting aspect is that this antibody has recognition properties that mimic the clinical severity of the disease. At least in HD, the recognition signal appears to increase over the whole length range of polyglutamine tracts, with no obvious threshold effect. The strong length dependence suggests that the antibody detects with high affinity a unique conformation that requires a minimum length of polyglutamine and that is stabilized by further increase in length. A structural model based on the analysis of small polyglutamine polypeptides did not predict such strong size dependence²⁶, and structural analysis of very long polyglutamine stretches is thus warranted. However, multiple bivalent bindings of the antibody on the homopolymeric epitopes might also account for part of our observations. Other proteins in neurons may have recognition properties similar to that of the 1C2 antibody, and such conformation-dependent interactions could contribute to pathogenesis. If so, it might be possible to design therapeutic molecules that interfere with the abnormal epitope, and even to target neurons of suitable animal models with a single-chain antibody derived from 1C2. □

Received 15 August; accepted 16 October 1995.

- La Spada, A. R., Wilson, E. M., Lubahn, D. B., Harding, A. E. & Fischbeck, K. H. *Nature* **352**, 77–79 (1991).
- The Huntington's Disease Collaborative Research Group *Cell* **72**, 971–983 (1993).
- Orr, H. T. et al. *Nature Genet.* **4**, 221–226 (1993).
- Koide, R. et al. *Nature Genet.* **6**, 9–13 (1994).
- Nagafuchi, S. et al. *Nature Genet.* **6**, 14–18 (1994).
- Kawaguchi, Y. et al. *Nature Genet.* **8**, 221–228 (1995).
- Housman, D. *Nature Genet.* **10**, 3–4 (1995).
- Albin, R. L. & Tagle, D. A. *Trends Neurosci.* **18**, 11–14 (1995).
- Benomar, A. et al. *Nature Genet.* **10**, 84–88 (1995).
- Gouw, L. G. et al. *Nature Genet.* **10**, 89–93 (1995).
- Gispert, S. et al. *Nature Genet.* **4**, 295–299 (1993).
- Belal, S. et al. *Neurology* **44**, 1423–1426 (1994).
- Pulst, S.-M., Nechiporuk, A. & Starkman, S. *Nature Genet.* **5**, 8–10 (1993).
- Gerber, H. P. et al. *Science* **263**, 808–811 (1994).
- Imbert, G., Trotter, Y., Beckmann, J. & Mandel, J.-L. *Genomics* **21**, 667–668 (1994).
- Gostout, B., Liu, Q. & Sommer, S. S. *Am. J. hum. Genet.* **52**, 1182–1190 (1995).
- Lescure, A. et al. *EMBO J.* **13**, 1166–1175 (1994).
- Trotter, Y. et al. *Nature Genet.* **10**, 104–110 (1995).
- Dürr, A. et al. *Brain* (in the press).
- Gardner, K. et al. *Neurology* **44**, A361 (1994) (Abstract).
- Ranum, L. P. W., Schut, L. J., Lundgren, J. K., Orr, H. T. & Livingston, D. M. *Nature Genet.* **8**, 280–284 (1994).
- Gispert, S. et al. *Am. J. hum. Genet.* **56**, 183–187 (1995).
- Hazan, J. et al. *Hum. molec. Genet.* **3**, 1569–1573 (1994).
- Hazan, J. et al. *Nature Genet.* **5**, 163–167 (1993).
- Ross, C. A., McInnis, M. G., Margolis, R. L. & Li, S.-H. *Trends Neurosci.* **16**, 254–260 (1993).

26. Stott, K., Blackburn, J. M., Butler, P. J. G. & Perutz, M. *Proc. natn. Acad. Sci. U.S.A.* **92**, 6509–6513 (1995).
 27. Trotter, Y., Biancalana, V. & Mandel, J.-L. *J. med. Genet.* **31**, 377–382 (1994).
 28. Servadio, A. et al. *Nature Genet.* **10**, 94–98 (1995).
 29. Sittler, A., Devys, D., Weber, C. & Mandel, J.-L. *Hum. molec. Genet.* (in the press).
 30. Lee, K. A. W., Bindereif, A. & Green, M. R. *Gene analyt. Techn.* **5**, 22–31 (1988).
 31. Lezin, A. et al. *Hum. Genet.* (in the press).

ACKNOWLEDGEMENTS. We thank E. Hirsch, C. Penet, N. Abbas, A. Benomar, A. Dürr, O. Didier-jean, A. Staub, S. Vicaire, N. Jung and V. Schultz for their help and contribution to this work; H. Cann for access to CEPH cell lines; and D. Rubinstein for the size of HD alleles in CEPH cell lines. This work was supported by grants from AFM, CNRS, INSERM, GREG Verum foundation and CHRU of Strasbourg. Y.T. was supported by a fellowship from Medical Research Council of Canada, and G.I. by MRT (Ecole Polytechnique).

Renal abnormalities and an altered inflammatory response in mice lacking cyclooxygenase II

Joseph E. Dinchuk^{*†}, Bruce D. Car, Richard J. Focht^{*}, Jennifer J. Johnston, Bruce D. Jaffee, Maryanne B. Covington, Nancy R. Contel, Vicki M. Eng[†], Robert J. Collins, Philip M. Czerniak, Stewart A. Gorry & James M. Trzaskos[†]

The DuPont Merck Pharmaceutical Company, Wilmington, Delaware 19880-0400, USA

^{*}The DuPont Merck Pharmaceutical Company, Glenolden, Pennsylvania 19036-2307, USA

[†]University Laboratory Animal Resources, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6021, USA

PROSTAGLANDINS have wide-ranging effects in the body and are thought to be important mediators of inflammation. Cyclooxygenase (COX) plays a key regulatory role in prostaglandin synthesis, and occurs in both constitutive (COX-1) and inducible (COX-2) isoforms^{1,2}. COX-1 is thought to provide cytoprotective effects³, whereas COX-2 is both inducible and the major isoform of inflammatory cells⁴. Reduction of prostaglandin production by inhibition of cyclooxygenases appears to be the main mechanism of action of most non-steroidal anti-inflammatory drugs (NSAIDs)⁵. Here we present an animal model of COX-2 deficiency that was generated by gene targeting. Defects in null mice correlating with reduced viability included renal alterations, characteristic of renal dysplasia (100% penetrance), and cardiac fibrosis (50% penetrance). Female *Cox-2*^{-/-} mice were infertile. COX-2 deficiency failed to alter inflammatory responses in several standard models, but striking mitigation of endotoxin-induced hepatocellular cytotoxicity was observed.

To investigate the roles of COX-2 in development and disease, stem cell technology was used to generate *Cox-2*^{-/-} mice (Fig. 1). The transcription and translation start sites of the COX-2 gene were removed in targeted cells. Chimaeric mice containing the targeted deletion were bred to produce both heterozygous and homozygous animals for the deletion. Analyses of cells and tissues showed that the knockout effectively eliminated expression of COX-2 message from alleles carrying the knockout (Fig. 2). Breeding yielded approximately 35% of the expected number of null animals. Neonatal deaths accounted for essentially all of this decrease, indicating that COX-2 deficiency was not lethal *in utero*. The average lifespan of null animals surviving to weaning was approximately 3.5 months, with relatively few null animals living beyond 6 months of age.

Histological examination of asymptomatic, adult *Cox-2*^{-/-} mice revealed tissue abnormalities restricted to the kidneys, heart

and ovaries (Fig. 3). Lesions were absent from heterozygote and wild-type animals. Haematological abnormalities were not observed in *Cox-2*^{-/-} mice, but adult *Cox-2*^{-/-} mice succumbed to chronic renal failure of developmental origin. The mammalian kidney contains both constitutive and inducible isoforms of cyclooxygenase^{6,7}, and the nephrotoxicity associated with NSAID administration could feasibly be the result of inhibition of either or both forms. Renal developmental abnormalities observed in *Cox-2*^{-/-} mice are not related to those of 'analgesic nephropathy'⁸, in which striking renal papillary necrosis precedes the development of a chronic tubulo-interstitial disease. Although kidneys from E14 *Cox-2*^{-/-} fetuses appeared normal, the kidneys in neonatal null animals were

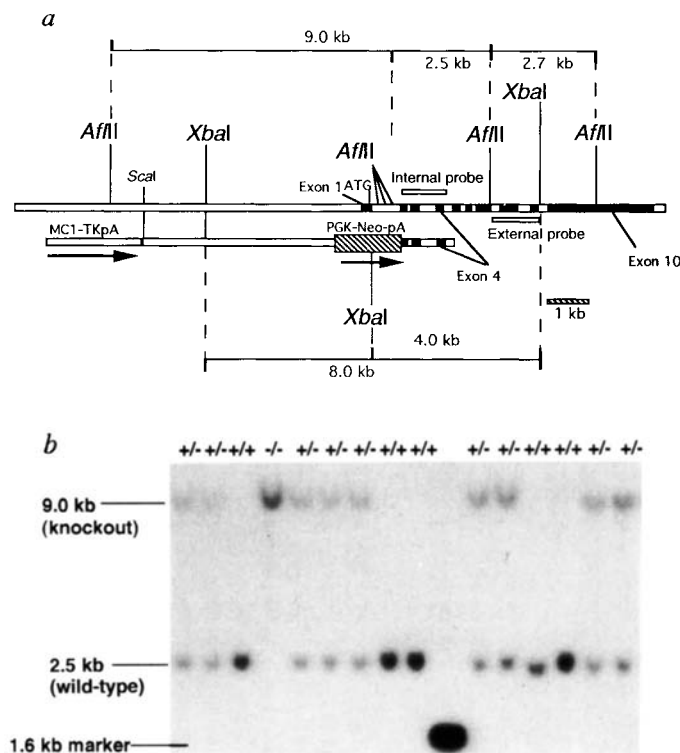


FIG. 1 Targeting of the COX-2 gene in embryonic stem (ES) cells and detection of the disruption in offspring. a, The targeting vector is shown next to the endogenous COX-2 gene. Crossover events will result in the incorporation of the 1.6-kb PGK-neo cassette (containing its own polyadenylation site) in place of a 1.8-kb EcoRV genomic fragment housing exon 1 and surrounding sequences. The neo cassette introduces an XbaI site and removes 3 endogenous AflII sites. With an internal or external probe, an XbaI digest yields a 4.0-kb fragment in recombinants and an 8.0-kb fragment in wild-type alleles. An internal probe and AflII digestion effectively tests for homologous recombination at both the 5' and 3' ends of the targeted gene simultaneously with the much larger (9.0 kb versus 2.5 kb) fragment appearing in targeted ES cells (data not shown). b, Mouse tail genomic DNA isolated from heterozygous crossing of offspring from a chimaeric founder was digested with AflII and analysed by Southern blotting using the internal probe. Wild-type (+/+), heterozygous (+/-) and homozygous (-/-) knockout offspring are indicated.

METHODS. The targeting construct was derived from a clone isolated from a 129/Sv genomic library (Stratagene). An MC1-tk cassette was placed at the 5' end of the targeting construct. The targeting vector was linearized at its 3' end and purified before introduction into AB2.1 ES cells by electroporation. A positive/negative selection strategy²⁷ using G418 and FIAU (1-(2'-deoxy-2-fluor α -D-arabinofuranosyl)-5-iodouracil) was used to select for recombinant ES clones. Enrichment in G418 and FIAU was 7.5-fold compared to G418 alone. Recombinant clones were injected into C57BL/6 blastocysts to generate chimaeras as described previously²⁸.

[†]To whom correspondence should be addressed.

TABLE 1 Inflammatory responses of COX-2-deficient mice

	Cox-2 +/+	Cox-2 +/-	Cox-2 -/-
Carrageenan-induced paw oedema* (Δ paw oedema, ml):			
129/Sv	0.15 $\pm$ 0.02 (n=13)	NT	0.12 $\pm$ 0.04 (n=3)
C57 X129	0.19 $\pm$ 0.06 (n=3)	0.21 $\pm$ 0.05 (n=4)	0.20 $\pm$ 0.03 (n=17)
TPA-induced ear oedema† (Δ ear swelling, mg):			
129/Sv	8.0 $\pm$ 1.8 (n=7)	NT	7.6 $\pm$ 2.3 (n=3)
Arachidonic-acid-induced ear oedema‡ (Δ ear swelling, units):			
C57 X 129	112.0 $\pm$ 5.7 (n=14)	117.0 $\pm$ 4.9 (n=7)	114.0 $\pm$ 10.9 (n=4)
Hepatocyte toxicity by LPS/D-galN-treatment§			
LPS/D-gal			
TNF (ng ml ⁻¹)	NT	7.3 $\pm$ 2.4	5.0 $\pm$ 1.0
ALT (IU)	NT	1,510 $\pm$ 930	300 $\pm$ 110
AST (IU)	NT	1,070 $\pm$ 480	270 $\pm$ 140
D-gal			
TNF (ng ml ⁻¹)	NT	0.0 $\pm$ 0.0	0.0 $\pm$ 0.0
ALT (IU)	NT	140 $\pm$ 20	160 $\pm$ 20
AST (IU)	NT	110 $\pm$ 50	90 $\pm$ 50

* Mice are injected in a rear paw with 0.05 ml carrageenan in 0.9% saline. Paw volumes were measured 3 h later (ml) ($\pm$ s.e.m.). No differences were noted between groups, even after extending measurements to 4 days (not shown).

† Tetradecanoyl phorbol acetate (1 μ g) in acetone is applied to the ears of the mice; 4 h later, 6 mm ear punches are taken and weighed (mg) ($\pm$ s.e.m.).

‡ Arachidonic acid (1 mg) in acetone is applied to the ears of the mice; 1 h later, ear swelling is measured in units of 10⁻⁴ inches ($\pm$ s.e.m.).

§ D-galactosamine (D-galN) (20 mg per mouse) and LPS (200 mg per kg; O111:B4 from *Escherichia coli*) given by intraperitoneal injection. Serum for TNF- α analysis (ELISA from Genzyme) was taken 1 h post-treatment and for serum aminotransferase (ALT) and aspartate aminotransferase (AST) 6 h post-treatment. Data represent mean $\pm$ s.e.m. and are typical for 2 experiments; n = 4–5 per group. Methods have been described previously²⁶.

severely underdeveloped, having few functional nephrons and an abundance of undeveloped mesenchymal tissue (Fig. 3). Mild to moderate renal lesions characterized by an overabundance of immature small glomeruli located subcapsularly and multiple foci of dysplastic tubules were present in all adult *Cox-2*^{-/-} mice examined (Fig. 3f). Corticomedullary microcysts and mild medullary hypoplasia or atrophy were also seen in these mice. *Cox-2*^{-/-} mice had increased blood urea nitrogen and serum creatinine (data not shown), consistent with renal histological abnormalities. These alterations, and the predisposition for development of secondary pyelonephritis (not shown), are consistent with renal dysplasia, suggesting that uterine bud induction of metanephric mesenchyme (embryonic day 15–18) into the normal glomerular and tubular structures of the mature nephron is partly dependent on COX-2 induction. A variety of genes expressed during metanephric differentiation have recently been identified^{9,10}. Mice deficient in Wnt-4 and those with deregulated expression of Pax-2 also display abnormalities in renal differentiation^{10,11}. Related, although milder, abnormalities in the kidneys of *Cox-2*^{-/-} mice suggest that these phenomena may be related.

Diffuse myocardial fibrosis involving both right and left ventricles with mild cardiac myofibre drop out occurred with varying severity in half of the adult null mice (Fig. 3g, h). Cardiac fibrosis, limited to mice deficient in COX-2, suggests that COX-2 deficiency might be involved in cardiomyopathy. There was no relationship between severity of renal and cardiac abnormalities.

Cox-2^{-/-} females were largely infertile, with only a few births recorded despite numerous matings. Ovaries of null mice

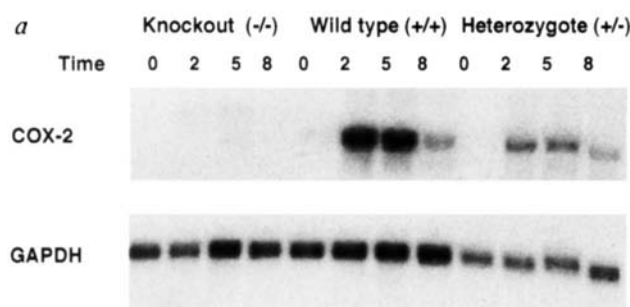
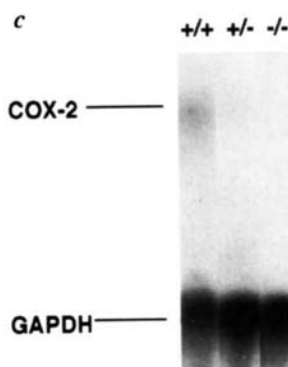
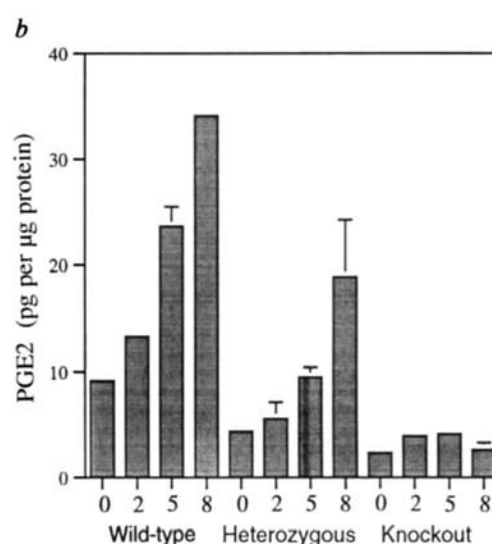


FIG. 2 COX-2 messenger RNA and PGE₂ production in primary embryonic fibroblasts (PEFs) and COX-2 mRNA from whole brain of identified mice. **a**, Northern blot analysis shows no normal COX-2 message in *Cox-2*^{-/-} fibroblasts, and less message in heterozygous fibroblasts. A GAPDH probe was used as a control for RNA loading/quality. **b**, The amount of COX-2 message correlates well with the level of PGE₂ produced in stimulated fibroblasts. A low level of PGE₂ is observed in null fibroblasts and may represent COX-1-derived prostaglandin. **c**, Northern blot analysis of whole brain from unstimulated, adult mice.

METHODS. Primary embryonic fibroblasts (PEFs) were obtained and cultured from 13-day-old embryos (heterozygous mating) as described previously²⁹. PEFs were analysed by Southern blotting and studied between passages 2 and 5. For the northern blot analysis, PEFs were plated in DMEM with 10% FBS at 2.8×10^5 cells per 100-mm dish. At 18 h, cells were washed with PBS and serum-starved (0.5% FBS) for 36 h. Cells were stimulated with 20% serum and then collected after 0, 2, 5 and 8 h. RNA (10 μ g per lane) was electrophoresed through a formaldehyde gel and transferred by blotting onto nitrocellulose or nylon. For whole brain, 15 μ g RNA was loaded per lane. The probe consisted of human COX-2 complementary DNA from nucleotides 98–1,926 (ref. 30). A rat GAPDH cDNA probe was used as a control for loading. For the PGE₂ assay, duplicate PEF cultures were prepared as for northern analysis. After 0, 2, 5 and 8 h, cells were pulsed with 100 μ M arachidonic acid in 1 ml DMEM with 0.5% FBS for 10 min. Cells were washed and disrupted in cell lysis buffer for protein determinations. ELISA was performed using the TiterZyme PGE₂ EIA kit (PerSeptive Diagnostics).



were small owing to a virtual absence of corpora lutea (Fig. 3*i,j*), but ovarian follicular development was apparently normal. These data support the concept that inhibition of COX-2 prevents ovulation¹²⁻¹⁴.

The concept that isoform-specific inhibitors might have advantages over current NSAIDs arose from the observation that the primary (and inducible) form of COX in inflammatory cells is COX-2. Newer NSAIDs, such as NS-398 and Dup-697, selectively inhibit COX-2 (refs 15, 16). The full potential of this new class of drugs will only be realized when it has been demonstrated that COX-2 inhibition is anti-inflammatory and has no significant side effects. Gastric and renal side effects are common with high-dose NSAID therapy. In animal models, COX-2-specific NSAIDs reduce inflammation without these side effects^{17,18}. The results of our studies support this latter observation.

COX-2-deficient mice and controls responded similarly when evaluated in carrageenan-induced paw oedema, TPA-induced

ear oedema, or arachidonic acid-induced oedema (Table 1). Although the arachidonic-acid ear-painting oedema model is likely to be wholly dependent on COX-1, several studies, including reports of a knockout of the 5-lipoxygenase gene, suggest a role for COX-2 in several animal models of inflammation induced by carrageenan and phorbol esters^{17,19-22}. When COX-2-deficient mice and controls were sensitized to tumour necrosis factor (TNF)- α -induced hepatocellular cytotoxicity with D-galactosamine, both groups produced similar levels of TNF- α in response to endotoxin (lipopolysaccharide) injection. In contrast, hepatocyte necrosis, as indicated by transaminase release, was clearly mitigated in mice deficient in COX-2 (Table 1). That macrophage priming for lipopolysaccharide (LPS)-induced TNF- α production was normal but transaminase release was markedly reduced suggests that prostaglandin synthesis by macrophages/Kupffer cells^{23,24} contributes significantly to cytotoxicity in this model. Conventional NSAIDs, which show selectivity for COX-1, do not protect normal mice from

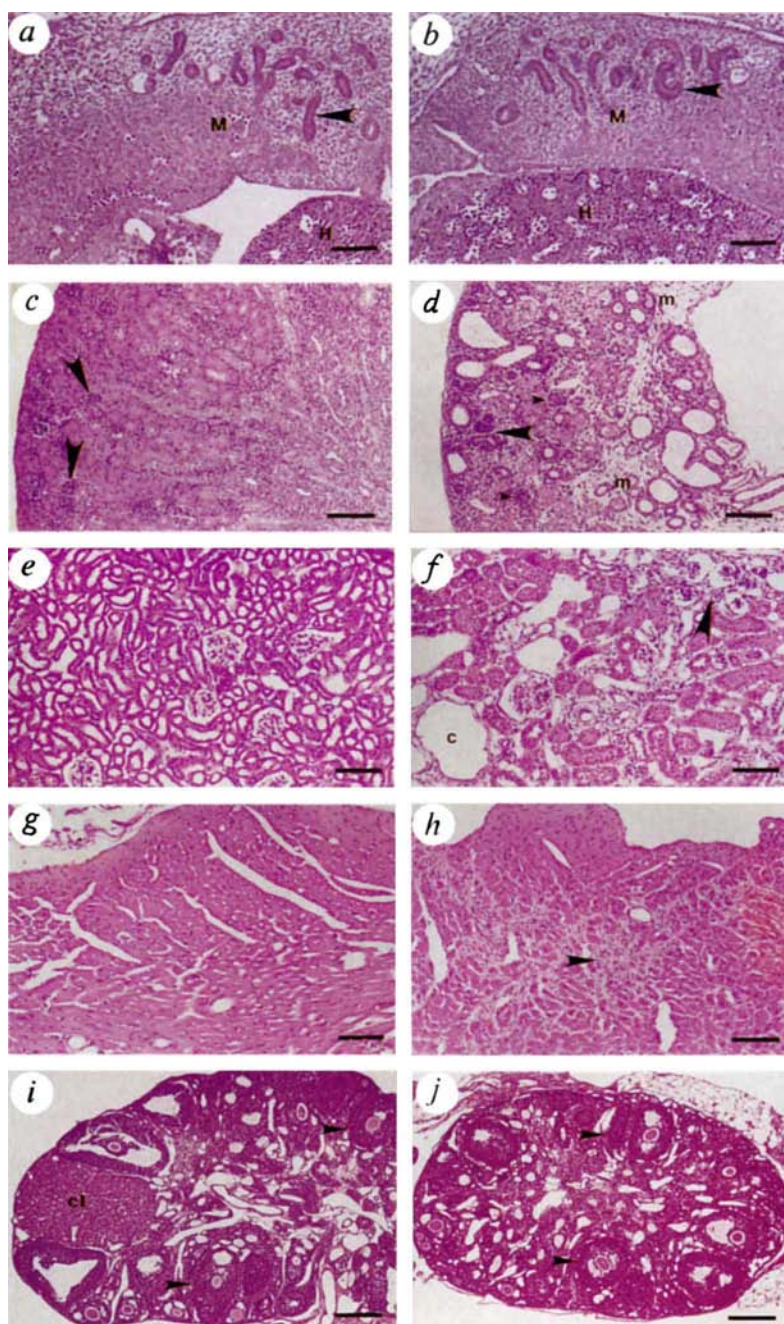


FIG. 3 *a*, Control 14-day-old fetal mouse metanephros (M) and liver (H) S-shaped bodies indicated by arrowhead (scale bar, 112 μ m); *b*, COX-2 knockout kidney showing similar degree of development; *c*, control 2-day-old mouse kidney, capsular border to the left, normal glomeruli indicated by arrowheads (scale bar, 112 μ m); *d*, COX-2 knockout mouse kidney, capsular border to the left, severe corticomedullary atrophy, S-shaped bodies (large arrowhead), abundant undifferentiated mesenchyme (m), poor collecting duct development with microcyst formation (right) (scale bar, 112 μ m); *e*, control adult outer renal cortex, renal capsule (c) (scale bar, 112 μ m); *f*, immature glomeruli (arrowhead), microcysts (m) (scale bar, 112 μ m); *g*, heart, left ventricular myocardium, free wall ventricular lumen top, control adult mouse (scale bar, 112 μ m); *h*, heart COX-2 knockout adult with diffuse myocardial fibrosis (arrowhead) (scale bar, 112 μ m); *i*, control adult ovary, corpus luteum (cl), ovarian follicles (small arrowheads; scale bar, 280 μ m); *j*, COX-2 knockout adult ovary (scale bar, 280 μ m); note absence of corpora lutea, and smaller size.

METHODS. Six mice from each group (wild-type, heterozygous and knockout) were anaesthetized in methophane. After weights of spleen, liver, lungs, heart, kidneys, thymus and whole brain were taken for each of the 18 animals the tissues were immediately divided. Adult tissues were perfused with 5% glutaraldehyde/4% paraformaldehyde and post-fixed in the same fixative. Mouse pup tissues were fixed in 4% formaldehyde/phosphate buffered saline. All tissues were sectioned at 5 μ m, stained with haematoxylin-eosin and photographed with a Nikon Microphot SA.

hepatocellular toxicity in this model²⁵. These data suggest that COX-2 inhibition mitigates LPS-induced hepatocellular injury and that inflammatory responses in the carrageenan and phorbol ester models are COX-2 independent. □

Received 10 May; accepted 27 October 1995.

- Kujubu, D. S., Fletcher, B. S., Varnum, B. C., Lim, R. W. & Herschman, H. R. *J. biol. Chem.* **266**, 12866–12872 (1991).
- Xie, W., Chipman, J. G., Robertson, D. L., Erikson, R. L. & Simmons, D. L. *Proc. natn. Acad. Sci. U.S.A.* **88**, 2692–2696 (1991).
- Mitchell, J. A., Akaraseenont, P., Thiemermann, C., Flower, R. J. & Vane, J. R. *Proc. natn. Acad. Sci. U.S.A.* **90**, 11693–11697 (1994).
- DeWitt, D. L., Meade, E. A. & Smith, W. L. *Am. J. Med. (suppl.)* **95**, 40S–44S (1993).
- Abramson, S. B. & Weissman, G. *Arthritis Rheum.* **32**, 1–9 (1989).
- Harris, R. C. et al. *J. clin. Invest.* **94**, 2504–2510 (1994).
- Martin, M. et al. *Kidney Int.* **45**, 150–158 (1993).
- Kinkaid-Smith, P. & Nanra, R. S. in *Diseases of the Kidney* 5th edn (eds Schrier, R. W. & Gottschalk, C. W.) 1099–1130 (Little, Brown and Co., Boston, MA, 1993).
- Pachnis, V., Mankoo, B. & Costantini, F. *Development* **119**, 1005–1097 (1993).
- Dressler, G. R. et al. *Nature* **362**, 65–67 (1993).
- Stark, K. S., Vainio, S., Vassilev, G. & McMahon, A. P. *Nature* **372**, 679–683 (1994).
- Sirois, J., Simmons, D. L. & Richards, J. S. *J. biol. Chem.* **267**, 11586–11592 (1992).

- Sirois, J., Levy, L. O., Simmons, D. L. & Richards, J. S. *J. biol. Chem.* **268**, 12199–12206 (1993).
- Wong, W. Y. L. & Richards, J. S. *Endocrinology* **130**, 3512–3521 (1992).
- Gans, K. R. et al. *J. Pharmac. exp. Ther.* **254**, 180–187 (1990).
- Gierse, J. K. et al. *Biochem. J.* **305**, 479–484 (1995).
- Seibert, K. et al. *Proc. natn. Acad. Sci. U.S.A.* **91**, 12013–12017 (1994).
- Masferrer, J. L. et al. *Proc. natn. Acad. Sci. U.S.A.* **91**, 3228–3232 (1994).
- Calhoun, W., Chang, J. & Carlson, R. P. *Agents & Actions* **21**, 306–309 (1987).
- Young, J. M. et al. *J. invest. Derm.* **82**, 367–371 (1984).
- Chen, X.-S., Sheller, J. R., Johnson, E. N. & Funk, C. D. *Nature* **372**, 179–182 (1994).
- Goulet, J. L. et al. *Proc. natn. Acad. Sci. U.S.A.* **91**, 12852–12856 (1994).
- Peters, T. et al. *Eur. J. Biochem.* **191**, 583–589 (1990).
- Ariasnegrete, S. et al. *Biochem. biophys. Res. Commun.* **208**, 582–589 (1995).
- Tiegs, G. & Wendel, A. *Biochem. Pharmac.* **37**, 2569–2573 (1988).
- Car, B. D. et al. *J. exp. Med.* **179**, 1437–1444 (1994).
- Mansour, S. L., Thomas, K. R. & Capecci, M. R. *Nature* **336**, 348–352 (1988).
- Ramirez-Solis, R., Davis, A. C. & Bradley, A. in *Guide to Techniques in Mouse Development* (eds Wassarman, P. M. & DePamphilis, M. L.) 855–878 (Academic, New York, 1993).
- Robertson, E. J. in *Teratocarcinomas and Embryonic Stem Cells. A Practical Approach* (ed. Robertson, E. J.) 71–112 (IRL, Oxford, 1987).
- Hia, T. & Neilson, K. *Proc. natn. Acad. Sci. U.S.A.* **89**, 7384–7388 (1992).

ACKNOWLEDGEMENTS. We thank G. Rule, N. Garcia, T. Cook and L. Smith for monitoring animal health; R. Senick for caesarean delivery of mice; and J. Jonak for helping establish the embryonic stem cell laboratory. Cox-2-targeted mice are available from the Induced Mutant Resource of The Jackson Laboratory under strain designation Pigs 2.

Involvement of Ral GTPase in v-Src-induced phospholipase D activation

Hong Jiang*, Jing-Qing Luo*, Takeshi Urano†, Paul Frankel*, Zhimin Lu*, David A. Foster*‡ & Larry A. Feig†

* The Institute for Biomolecular Structure and Function, and † Department of Biological Sciences, The Hunter College of The City University of New York, 695 Park Avenue, New York, New York 10021, USA

‡ Department of Biochemistry, Tufts University School of Medicine, Boston, Massachusetts 02111, USA

AN early response to the tyrosine kinase activity of v-Src is an increase in phospholipase D (PLD) activity¹, which leads to the generation of biologically active lipid second messengers, including phosphatidic acid, lysophosphatidic acid and diacylglycerol². We have recently demonstrated that v-Src-induced PLD activity is mediated by Ras³, although Ras involvement was indirect, requiring a cytosolic factor for PLD activation³. Ras interacts with^{4–6} and activates Ral-GDS¹³, the exchange factor responsible for the activation of Ral GTPases. Here we report that this newly identified Ras/Ral signalling pathway mediates PLD activation by v-Src. PLD activity could be precipitated from v-Src-transformed cell lysates with immobilized RalA protein and with an anti-Ral antibody. A mutation to the region of RalA analogous to the 'effector domain' of Ras did not reduce the ability of RalA to complex with PLD, although deletion of a Ral-specific amino-terminal region did. Overexpression of RalA potentiated PLD activation by v-Src, and expression of dominant negative RalA mutants inhibited both v-Src- and v-Ras-induced PLD activity. Thus RalA is involved in the tyrosine kinase activation of PLD through its unique N terminus, and that PLD is a downstream target of a Ras/Ral GTPase cascade.

We recently reported that, although the activation of phospholipase D (PLD) by v-Src is mediated by Ras³, two reported Ras effector molecules, Raf-1 (refs 7, 8) and phosphatidylinositol-3-kinase⁹, were apparently not involved³. Our recent observation that Ras also activates Ral-GDS¹³ suggested that the activation of PLD by v-Src could be through Ral GTPase. We therefore examined whether glutathione-S-transferase (GST)-RalA fusion proteins (immobilized Ral) could associate with

PLD activity in detergent lysates of v-Src-transformed and parental BALB/c 3T3 cells. Both GTP- and GDP-loaded RalA precipitated PLD activity from both cell types (Fig. 1a, b). Although GTP-bound RalA was more efficient than GDP-bound RalA in binding PLD, the difference was substantially less than that reported for the binding of both Ras and Ral to other downstream effector molecules^{7–10}. The effect was specific for Ral, as GST-Ras was unable to precipitate PLD activity (Fig. 1b). To determine whether complexes between Ral and PLD exist *in vivo*, we investigated whether PLD activity could be precipitated by an anti-Ral antibody. An anti-Ral, but not an anti-Ras, antibody precipitated PLD activity from v-Src-transformed cell lysates (Fig. 1c). PLD activity could also be precipitated from parental BALB/c 3T3 cell lysates by the anti-Ral antibody, suggesting that an association between RalA and PLD exists before tyrosine kinase activation.

We next examined the ability of RalA mutants to associate with PLD. D49N RalA has an Asp to Asn substitution at position 49 of RalA, which is in a region of RalA analogous to the effector domain of Ras that associates with downstream effector molecules. The D49N mutation in RalA has previously been shown to prevent association between RalA and a different putative downstream target molecule, Ral-BP1 (ref. 10). Surprisingly, the association between RalA and PLD was actually enhanced by the D49N mutation (Fig. 1d). In contrast, deletion of 11 N-terminal residues (Δ N11) unique to Ral proteins strongly reduced the association between RalA and PLD (Fig. 1d). The Δ N11 RalA deletion did not significantly affect the association of RalA with Ral-BP1 (ref. 10; our unpublished results).

We next examined the effect of altering the activity of Ral in intact cells. Wild-type RalA and an activated RalA mutant (Q72L) were overexpressed in v-Src-transformed and parental NIH 3T3 cells. Both wild-type and activated RalA increased PLD activity in the v-Src-transformed cells, but had no effect upon the PLD activity in the parental cells (Fig. 2a). Purified GTP-bound RalA did not increase PLD activity in cell lysates (not shown). Thus, although RalA contributes to v-Src-induced PLD activation, it is not sufficient for PLD activation.

We next tested whether RalA is necessary for v-Src activation of PLD by examining the effect of a dominant negative RalA mutant (S28N) on v-Src-induced PLD activity. The S28N mutation is analogous to the dominant negative S17N mutant of Ras^{11,12}. Expression of S28N RalA in v-Src-transformed cells reduced PLD activity to the level observed in the parental cells or in v-Src-transformed cells expressing the dominant negative S17N Ras mutant (Fig. 2b). This effect was observed only in cells prelabelled with [³H]-myristate, not in cells prelabelled with

‡ To whom correspondence should be addressed.

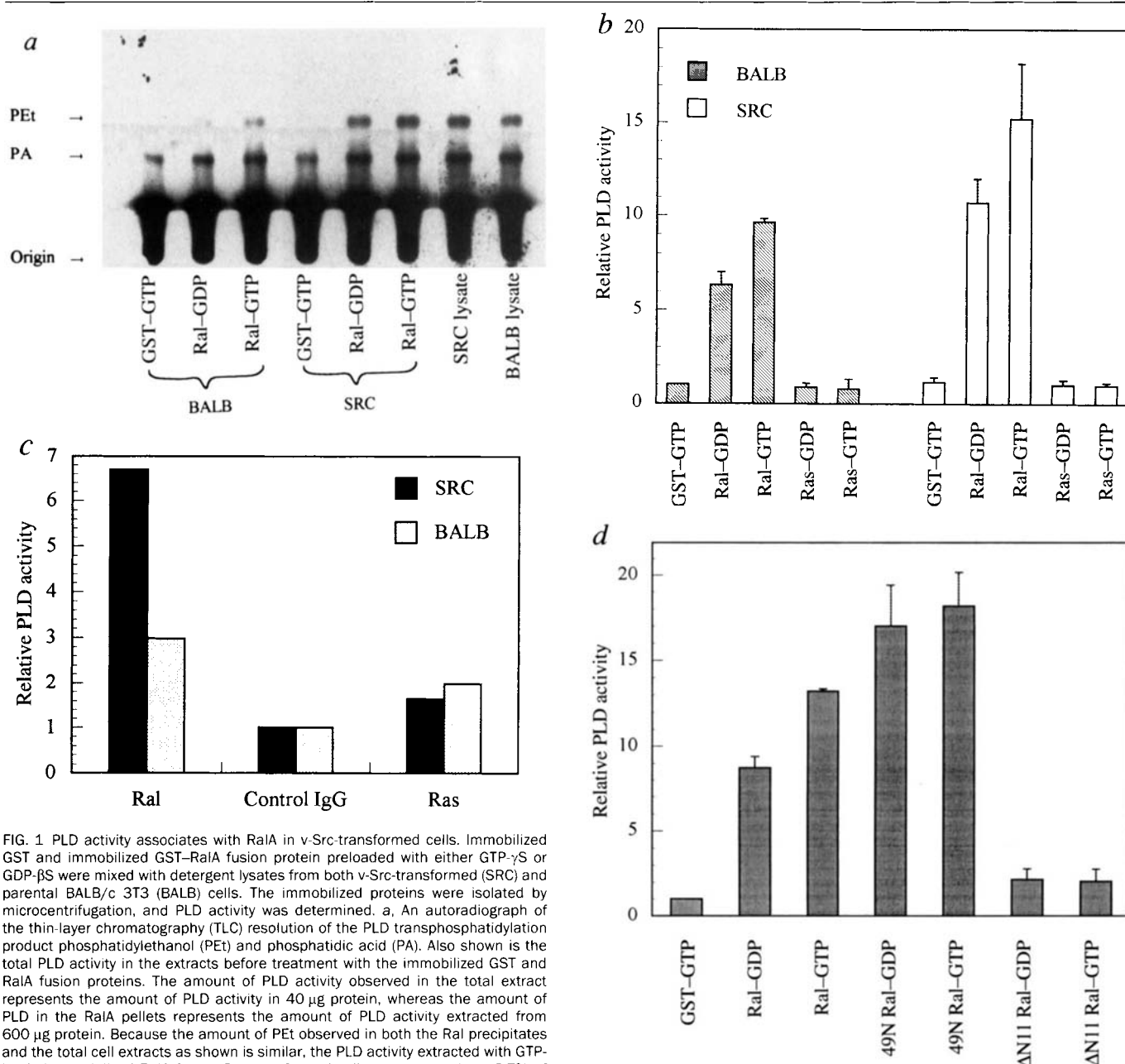
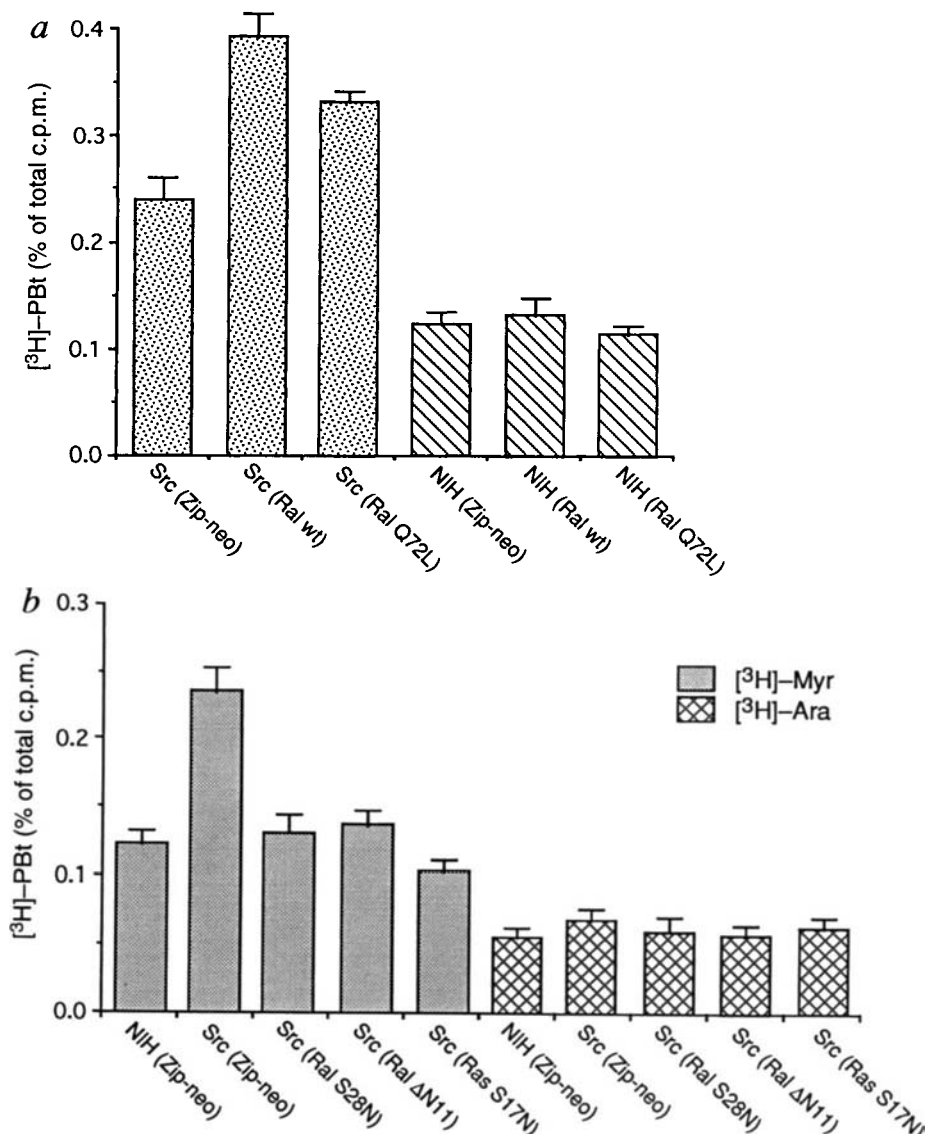


FIG. 1 PLD activity associates with RalA in v-Src-transformed cells. Immobilized GST and immobilized GST-RalA fusion protein preloaded with either GTP- γ S or GDP- β S were mixed with detergent lysates from both v-Src-transformed (SRC) and parental BALB/c 3T3 (BALB) cells. The immobilized proteins were isolated by microcentrifugation, and PLD activity was determined. **a**, An autoradiograph of the thin-layer chromatography (TLC) resolution of the PLD transphosphatidylation product phosphatidylethanol (PEt) and phosphatidic acid (PA). Also shown is the total PLD activity in the extracts before treatment with the immobilized GST and RalA fusion proteins. The amount of PLD activity observed in the total extract represents the amount of PLD activity in 40 μ g protein, whereas the amount of PLD in the RalA pellets represents the amount of PLD activity extracted from 600 μ g protein. Because the amount of PEt observed in both the Ral precipitates and the total cell extracts as shown is similar, the PLD activity extracted with GTP-loaded immobilized RalA from v-Src-transformed cells represents about 6.7% of total PLD. **b**, Immobilized GST, GST-RalA and GST-Ras fusion proteins preloaded with either GTP- γ S or GDP- β S as shown were mixed with detergent lysates of either v-Src-transformed or parental BALB/c 3T3 cells and the relative PLD activity in the pellets was determined. **c**, Detergent lysates from v-Src-transformed and parental BALB/c 3T3 cells were treated with IgG purified from a polyclonal anti-Ral antiserum¹³, monoclonal anti-Ras antibody 259 (ref. 23) and control rabbit IgG overnight at 4 °C. The immune complexes were recovered with protein A agarose as described previously³, and associated PLD activity was determined as in **a** and **b**. **d**, The ability of GTP- γ S- and GDP- β S-bound immobilized RalA mutants to associate with PLD was determined. PLD activity associated with wild-type RalA is included for comparison. D49N contains an amino-acid substitution of Asn for Asp at position 49; S28N contains a substitution of Asn for Ser at position 28; and ΔN11 has a deletion of 11 amino acids at the N terminus. In **b**, the values were normalized to the PLD activity associated with the GST control. The error bars reflect the range of the PLD activity relative to the GST controls for two independent experiments performed in duplicate exclusive to the representative experiment shown in **a**. In **c**, an autoradiograph of the TLC plate was scanned using a Molecular Dynamics densitometer. The relative PLD activity associated with the anti-Ral, anti-Ras and control IgG was determined by comparing the relative intensities of the phosphatidylethanol bands. The PLD activity was normalized to the PLD activity that was detected in the control IgG precipitates. The representative experiment shown is one of two performed, in which the results varied by less than 10%. In **d**, the data were generated as in **b** and then normalized to the PLD activity precipitated with GST alone. The error bars reflect the range of the PLD activity relative to the GST control for two independent experiments. **METHODS**. v-Src-transformed BALB/c 3T3 cells were maintained as described previously⁷. Cell cultures were grown to confluence, at which time the media was replaced with fresh media containing 0.5% newborn calf serum for 1 d. Cells were

treated with lysis buffer (25 mM HEPES, pH 7.2, 100 mM KCl, 10 mM NaCl, 0.5 mM EGTA, 0.5 mM EDTA, 1 mM DTT) containing Triton X-100 (1%). Protease inhibitors (12 μ g ml⁻¹ leupeptin, 20 μ g ml⁻¹ aprotinin, 0.1 mM phenylmethylsulfonyl fluoride) were included just before lysis (30 min at 4 °C). Lysates were scraped from the plates with a rubber policeman and centrifuged for 10 min at 1,500 r.p.m.; the supernatant was then centrifuged for 45 min at 30,000 r.p.m. and the supernatant was recovered and protein concentration determined using the Bio-Rad assay. Cell lysates (600 μ g protein) were treated with immobilized GST fusion proteins as described above. After 1.5 h at 4 °C, the GST fusion proteins were separated from the lysate by microcentrifugation for 30 s and washed three times with cold lysis buffer. A liposome-based *in vitro* PLD assay¹⁹ was used to determine the PLD activity associated with the Ral and Ras fusion proteins. PLD activity was determined by examining the ability to convert [³H]-PC in prepared liposomes to phosphatidylethanol in the presence of exogenously provided ethanol. Liposomes were prepared by mixing chloroform solutions of phospholipids containing 1 μ Ci per reaction of [³H]-PC (42 Ci mM⁻¹) and drying under a stream of nitrogen followed by resuspension with sonication for 10 min at 45 °C in lysis buffer. The liposomes consist of phosphatidylethanolamine/phosphatidylinositol-4,5-bisphosphate/PC in a molar ratio of 16/1.4/1 with PC suspended to a final concentration of 8.6 μ M. The inclusion of phosphatidylinositol-4,5-bisphosphate in the liposomes was essential¹⁹. The reaction buffer consisted of lysis buffer plus 5 mM MgCl₂, 0.16 mM CaCl₂, and 1% ethanol in a total volume of 0.2 ml. The reaction mixtures were incubated for 15 min and terminated by the addition of 2 ml CHCl₃:MeOH:HCl 1:1:0.006. Phase separation was achieved by adding 1 ml of 0.1 M HCl, 1 mM EGTA. The resolution of PLD products by TLC and quantification have been described previously³. The construction of the RalA mutants and the loading of immobilized Ras and Ral with guanine nucleotides were as described previously^{3,10}.

FIG. 2 The effect of wild-type and mutant RalA on PLD activity in NIH 3T3 cells. **a**, The effect of wild-type and an activated mutant of RalA (Q72L) on PLD activity in v-Src-transformed and parental NIH 3T3 cells was determined. v-Src-transformed and parental NIH 3T3 cells were stably transfected with vectors expressing wild-type and the Q72L RalA genes. The Q72L RalA mutant contains a mutation analogous to the activating Ras mutation at position 61. RalA proteins were expressed at approximately sixfold greater concentrations than endogenous RalA (not shown). The PLD activity in these cells was determined by measuring the level of phosphatidylbutanol (Pbt) generated in the presence of exogenously supplied butanol (1.0% for 30 min). The Ral protein expressed by Q72L is defective for GTPase activity, as is the homologous mutation in Ras at position 61 (ref. 24). **b**, The effect of potential dominant negative RalA mutants was compared with the effect of a dominant negative Ras mutant on PLD activity in v-Src-transformed NIH 3T3 cells. The v-Src-transformed NIH 3T3 cells were stably transfected with plasmids expressing the S28N and Δ N11 mutants of RalA and the dominant negative S17N Ras mutant¹¹. NIH 3T3 (NIH), v-Src-transformed NIH 3T3 cells (Src), and the v-Src-transformed cells expressing the RalA and Ras mutants were prelabelled with either [³H]-myristate or [³H]-arachidonate as shown for 4–6 h. [³H]-myristate is incorporated exclusively into PC, the substrate for v-Src-induced PLD, whereas [³H]-arachidonate is incorporated into phospholipids not utilized by the v-Src-induced PLD^{1,14}. The S28N mutant had no significant effect upon either Src protein or phosphotyrosine levels (data not shown). The data in **a** and **b** are the average of duplicates $\pm$ the range from representative experiments repeated at least three times. The data are presented as amount of radioactivity present in the transphosphatidylated product, phosphatidylbutanol, as a percentage of the total c.p.m. loaded onto the resolving TLC plates. Samples are normalized for total c.p.m. before loading. There were no significant differences in the amount of either [³H]-myristate or [³H]-arachidonate uptake per cell for any of the cell lines used. **METHODS.** v-Src-transformed NIH 3T3 cells were stably transfected with plasmids carrying the RalA and Ras genes using G418 selection, as described previously³. RalA mutant expression vectors were constructed as described in ref. 13. NIH 3T3 cells were used instead of BALB/c 3T3



[³H]-arachidonate. We have previously used this differential-labelling strategy to distinguish between PLD activity induced by v-Src and that induced by phorbol ester¹⁴. Consistent with phorbol ester-induced PLD activity having a different mechanism, the S28N mutant had no effect upon the PLD activity induced by phorbol esters (not shown). These data, and those in Fig. 2a, support the notion that RalA is important for the activation of PLD by v-Src. Further support was provided by studying v-Src-transformed cells expressing the N-terminal RalA deletion mutant (Δ N11) that precipitated PLD poorly *in vitro*. PLD activity in these cells was reduced almost to the level observed in the parental NIH 3T3 cells. Apparently the Δ N11 mutant that failed to bind PLD efficiently *in vitro* was not able to mediate PLD activation by v-Src *in vivo*, and competed with endogenous RalA in this signalling pathway.

Because v-Src-induced PLD activity is dependent upon Ras³, and Ras has been reported to induce an increase in PLD

cells because of a higher transfection efficiency, which allowed us to pool selected clones of transfected cells to avoid clonal variants. Increased expression of the RalA and Ras protein levels was verified by western blot analysis (data not shown). PLD activity was determined as described previously³.

activity¹⁵, we examined the effect of the S28N RalA mutant on PLD activity in cells transformed by Ras. Expression of the S28N mutant inhibited the Ras-induced increase in PLD activity (Fig. 3). The degree of inhibition correlated well with the level of S28N Ral expression in several independent clones.

The experiments described here indicate that RalA mediates the tyrosine kinase activation of PLD through an interaction with N-terminal sequences unique to Ral. A model of how RalA might function in this capacity is outlined in Fig. 4. Ras is activated by v-Src and then brings Ral-GDS to the plasma membrane. Ral-GDS then brings RalA and constitutively associated PLD into a tyrosine kinase/Ras signalling complex. Because activated RalA (Q72L) is unable to activate PLD activity in non-transformed NIH 3T3 cells, an additional signal induced by v-Src must be required for PLD activity to increase. The observation that Ras-transformed cells also display increased PLD activity suggests that this additional signal is downstream

FIG. 3 Ras-induced PLD activity is inhibited by a dominant negative RalA mutant. NIH 3T3 cells were transformed by a vector expressing v-Ras (pv-HaRas) (lanes B–H) or the parental vector (pZip-neoSV(X))²⁵ (lane A). The Ras-transformed cells were then co-transfected with the vector expressing the S28N RalA mutant and pCEP4 (Invitrogen) (lanes D–H), which expresses the resistance gene for hygromycin. The ratio of the S28N vector to pCEP4 was 10:1. Hygromycin selection was used because the Ras-transformed cells already expressed the G418 resistance gene present in the S28N vector. Five independent hygromycin-resistant clones were picked, expression of Ral protein (endogenous and S28N) was determined by western analysis, and the relative levels of Ral protein were determined using densitometer tracing of the autoradiogram (a). The relative levels of PLD activity (normalized to the level of PLD activity in NIH 3T3 cells) were then determined as in Fig. 2 using [³H]-myristate-labelled cells (b).

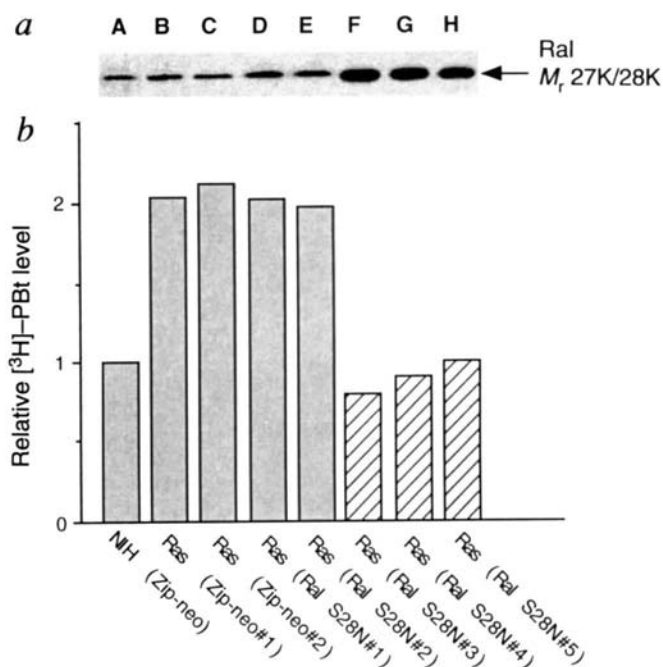
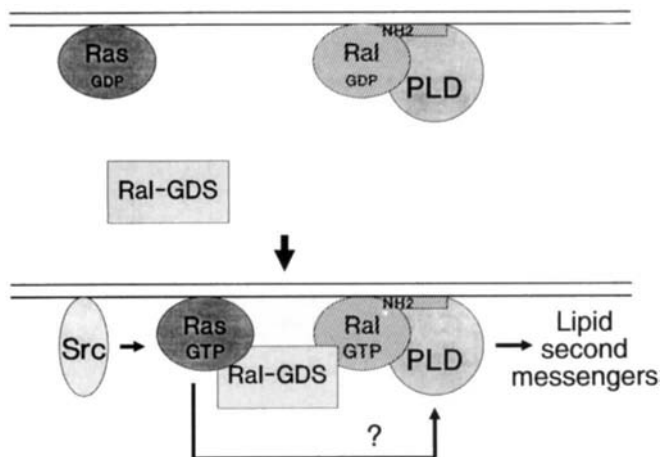


FIG. 4 Model for RalA involvement in v-Src-induced PLD activation. It is proposed that RalA, which is a membrane-bound protein, serves as an anchor for PLD through its amino terminus. Ral-GDS could then serve as a conduit for bringing PLD into a complex with activated RAS. An additional Ras downstream effector molecule(s) is postulated to account for the ability of activated Ras, but not activated Ral to increase PLD activity in non-transformed NIH 3T3 cells.



of Ras. Putative mediators of this signal are the Rho and ARF proteins, which are related to Ras and have been shown to stimulate PLD activity *in vitro*^{16–21}. There is genetic evidence linking RhoA and Ras²², but a connection between ARF and Ras remains to be demonstrated.

The significance of PLD activation by tyrosine kinases is unknown. However, PLD activity increases in response to all of the tyrosine kinases so far examined². The ability of PLD to generate multiple lipid second messengers by its primary metabolite phosphatidic acid suggests that PLD contributes significantly to cellular responses activated by Ras and mediated by Ras. In this regard, it is possible that PLD activation via RalA contributes to the ability of RalA to enhance both Ras- and Raf-induced cellular transformation¹³. □

Received 2 July; accepted 5 October 1995.

1. Song, J., Pfeffer, L. M. & Foster, D. A. *Molec. cell. Biol.* **11**, 4903–4908 (1991).
2. Foster, D. A. *Cell Signalling* **5**, 389–399 (1993).
3. Jiang, H., Lu, Z., Luo, J. Q., Wolfman, A. & Foster, D. A. *J. biol. Chem.* **270**, 6006–6009 (1995).
4. Kikuchi, A., Demo, S. D., Ye, Z., Chen, Y. & Williams, L. T. *Molec. cell. Biol.* **14**, 7483–7491 (1994).
5. Hofer, F., Fields, S., Schneider, C. & Martin, G. S. *Proc. natn. Acad. Sci. U.S.A.* **91**, 11089–11093 (1994).

6. Spaargaren, M. & Bischoff, J. R. *Proc. natn. Acad. Sci. U.S.A.* **91**, 12609–12613 (1994).
7. Moodie, S. A., Willumsen, B. M., Weber, M. J. & Wolfman, A. *Science* **260**, 1658–1661 (1993).
8. Vojtek, A. B., Holenberg, S. M. & Cooper, J. A. *Cell* **74**, 205–214 (1993).
9. Rodriguez-Viciana, P. et al. *Nature* **370**, 527–532 (1994).
10. Cantor, S., Urano, T. & Feig, L. A. *Molec. cell. Biol.* **15**, 4578–4584 (1995).
11. Feig, L. A. & Cooper, G. M. *Molec. cell. Biol.* **8**, 3235–3243 (1988).
12. Farnsworth, C. L. & Feig, L. A. *Molec. cell. Biol.* **11**, 4822–4829 (1991).
13. Urano, T., Emkey, R. & Feig, L. A. *EMBO J.* (in the press).
14. Song, J. & Foster, D. A. *Biochem. J.* **294**, 711–717 (1993).
15. Carnero, A., Cuadrado, A., del Peso, L. & Lacal, J. C. *Oncogene* **9**, 1387–1395 (1994).
16. Bowman, E. P., Uhlinger, D. J. & Lambeth, J. D. *J. biol. Chem.* **268**, 21509–21512 (1993).
17. Malcolm, K. C., Ross, A. H., Qui, R.-G., Symons, M. & Exton, J. E. *J. biol. Chem.* **269**, 25951–25954 (1994).
18. Siddiqui, A. R. et al. *J. biol. Chem.* **270**, 8466–8473 (1995).
19. Brown, H. A., Gutowski, S., Moomaw, C. R., Slaughter, C. & Strinweis, P. C. *Cell* **75**, 1137–1144 (1993).
20. Cockcroft, S. et al. *Science* **263**, 523–526 (1994).
21. Lambeth, J. D. et al. *J. biol. Chem.* **270**, 2431–2434 (1995).
22. Ridley, A. J. & Hall, A. *Cell* **70**, 389–399 (1992).
23. Furth, M. E., Davis, L. J., Fleurbaey, B. & Scolnick, E. M. *J. Virol.* **43**, 294–304 (1982).
24. Emkey, R., Friedman, S. & Feig, L. A. *J. biol. Chem.* **266**, 9703–9706 (1991).
25. Qureshi, S. A. et al. *J. biol. Chem.* **267**, 17635–17639 (1992).

ACKNOWLEDGEMENTS. H.J. and J.-Q.L. contributed equally to this work. We thank J. Brugge and R. Jove for the v-Src-transformed Balb/c 3T3 and NIH 3T3 cells, respectively; U. Rapp for the dominant negative Raf-1 mutant; S. Bychenok and M. Curto for comments on the manuscript; and X. Liu for helping to develop the *in vitro* PLD assays. This work was supported by the NIH, the Council for Tobacco Research, and the American Cancer Society (to D.A.F.); a Research Centers in Minority Institutions award from the Division of Research Resources, NIH; a Human Frontiers Science Program long term fellowship (to T.U.); and the NIH and an American Cancer Society faculty research award (to L.A.F.).

Conversion of thrombin into an anticoagulant by protein engineering

C. S. Gibbs, S. E. Coutré*, M. Tsiang, W.-X. Li, A. K. Jain, K. E. Dunn, V. S. Law, C. T. Mao, S. Y. Matsumura, S. J. Mejza, L. R. Paborsky & L. L. K. Leung*

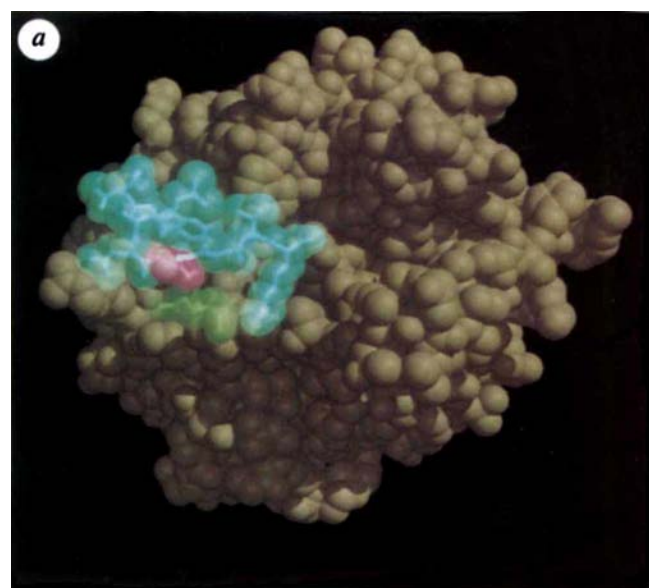
Gilead Sciences, Foster City, California 94404, USA

At sites of vascular injury, thrombin interacts with multiple procoagulant substrates¹⁻⁶ to mediate both fibrin clotting and platelet aggregation. But upon binding to thrombomodulin on the vascular endothelium, thrombin instead activates protein C, thereby functioning as an anticoagulant and attenuating clot formation⁷. Upon infusion *in vivo*, both the procoagulant and anticoagulant effects of thrombin were observed^{8,9}. Preliminary studies indicating that thrombin's protein C activating and fibrinogen clotting activities could be dissociated by mutagenesis¹⁰ suggested to us that a thrombin variant that lacked procoagulant activity while retaining anticoagulant function might be an attractive antithrombotic agent. Using protein engineering, we introduced a single substitution, E229A, that substantially shifted thrombin's specificity in favour of the anticoagulant substrate, protein C. In monkeys, this modified thrombin functioned as an endogenous protein C activator demonstrating dose-dependent, reversible anticoagulation without any indication of procoagulant activity. Notably, template bleeding times were not prolonged, suggesting a reduced potential for bleeding complications.

A library of 62 thrombin mutants¹¹, in which the solvent-exposed charged and polar amino acids were substituted with alanine, was screened for mutants that were defective in fibrinogen clotting yet retained their ability to activate protein C in the presence of thrombomodulin. One mutation, E229A, enhanced the ratio of protein C activation (in the presence of thrombomodulin) to fibrinogen clotting activity by 22-fold. This mutant was selected as a prototype protein C activator (PCA), overexpressed in BHK cells, purified to homogeneity and characterized *in vitro* (Table 1).

Recognition of the A α chain of fibrinogen by PCA, as determined by the catalytic efficiency of fibrinopeptide A (FPA) release, was diminished by 40-fold and this catalytic deficiency was reflected in an increase in the concentration required to achieve a clotting time of 60 s in human plasma (58 nM for PCA compared to 3.9 nM for wild-type thrombin). Similarly, the amount of thrombin required to achieve half-maximal platelet aggregation (EC₅₀) was increased 78-fold. In contrast, the affinity of PCA for thrombomodulin was unaltered and the catalytic efficiency of thrombomodulin-dependent protein C activation was substantially preserved (47.5% of wild-type thrombin). Therefore, the E229A mutation dramatically reduced the procoagulant activities of thrombin while anticoagulant activity in the presence of thrombomodulin was relatively unperturbed. In addition, PCA displayed resistance to antithrombin III, the principal physiological inhibitor of thrombin. The second-order rate constant for antithrombin III inhibition was decreased by sevenfold which may be correlated with the sixfold increase in the *in vitro* plasma half-life.

In the crystal structure of human thrombin complexed with FPA¹² (Fig. 1a), thrombin residue E229 is in van der Waals contact with the glycine residue at P5 in FPA. This interaction appears to be important as E229 is absolutely conserved among thrombins from nine vertebrate species¹³ and the P5 glycine is



b

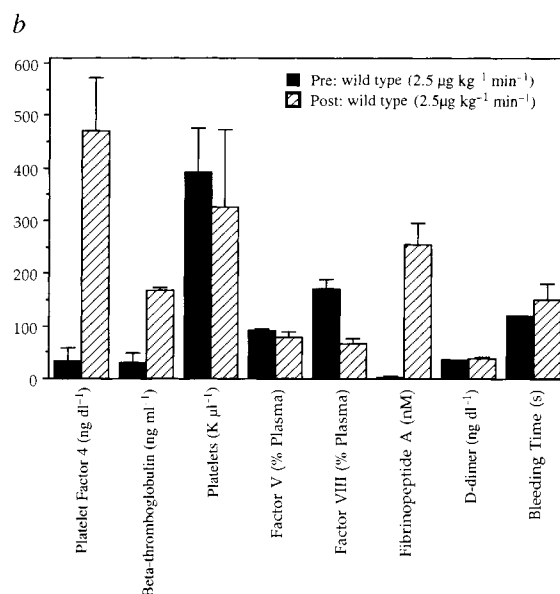
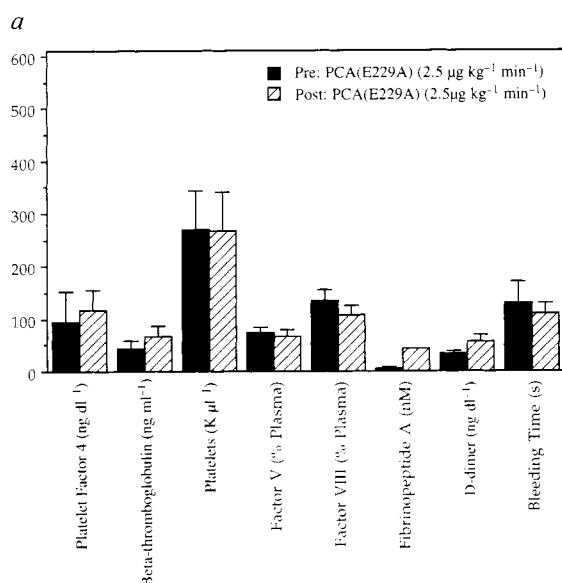
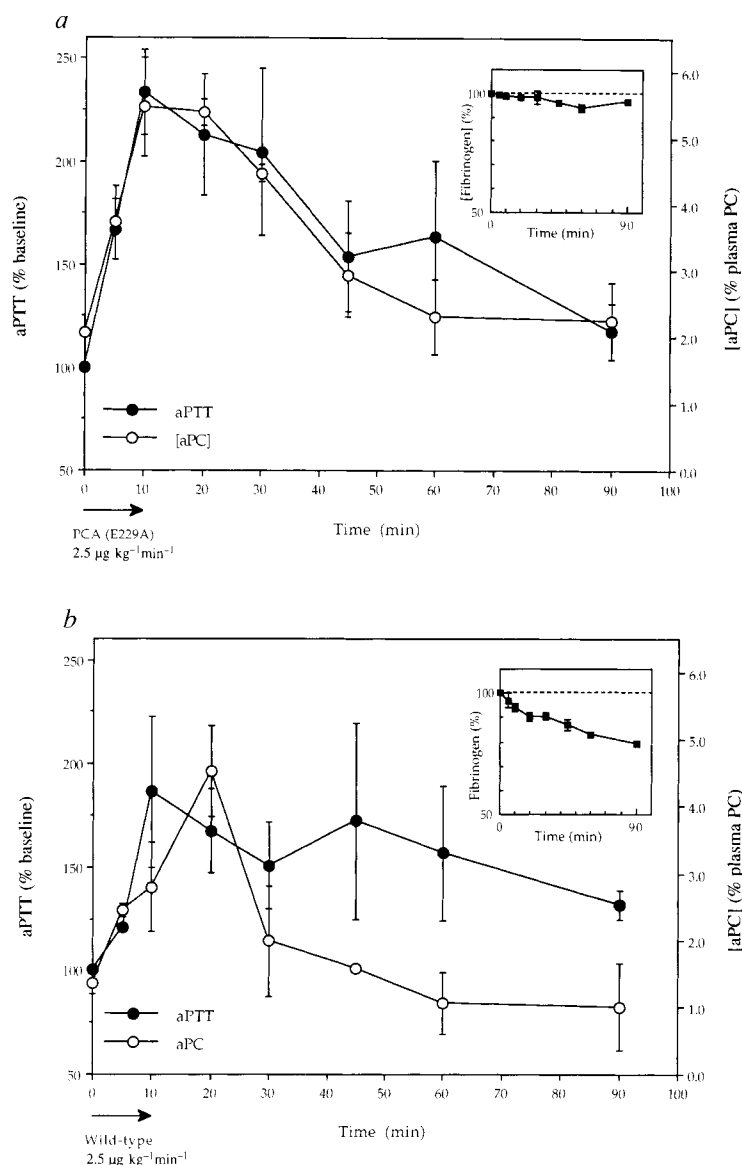
P	10	9	8	7	6	5	4	3	2	1	
	D	F	L	A	E	G	G	G	V	R	Human FPA
	D	F	L	T	E	G	G	G	V	R	Bovine FPA
	T	F	E	K	E	G	G	G	G	R	Chicken FPA
	E	F	I	A	E	G	G	G	V	R	Dog FPA
	S	F	Q	K	E	G	G	G	V	R	Duck FPA
	D	F	L	A	E	G	G	G	V	R	Japanese Macaque FPA
	E	F	I	E	A	G	G	D	I	R	Rat FPA
	D	F	L	A	E	G	G	G	V	R	Red Guenon FPA
	E	D	Q	E	D	Q	V	D	P	R	Human Protein C

FIG. 1 Structural basis for the involvement of thrombin residue E229 in the recognition of fibrinogen. a, Model of the interaction of FPA (blue) with the active site cleft of human thrombin based on the crystal structure of the complex¹². The glycine residue (magenta) at P5 in FPA is shown in van der Waals contact with E229 (green) in thrombin. Using chymotrypsinogen numbering of bovine trypsin, thrombin residue E229 = E217. b, Alignment of residues (P1–P10) proximal to the cleavage site in FPA from eight higher vertebrate species, in the NBRF-PIR and SwissProt data bases, with the activation peptide of human protein C. FPA binds thrombin in a hooked conformation (a) because of a type II β -turn that allows the phenylalanine residue at P9 to interact with the aryl binding site near the P2 binding site. This reverse turn is possible because the glycine residue at P5 assumes a conformation that would cause strain with any other residue¹². The non-conservation of residues flanking the cleavage site in protein C, including the P5 glycine, suggests that the protein C activation peptide may interact with the thrombin active site in an extended conformation that is less affected by the substitution of E229. Alternatively, the proposed conformational change associated with thrombomodulin binding⁷ may diminish the role of E229 in substrate recognition. However, in the absence of thrombomodulin, protein C activation by PCA and wild-type thrombin was decreased in parallel, suggesting the role of E229 is unaffected by thrombomodulin. Recently, the binding of Na⁺ to thrombin was shown to mediate an allosteric transition to a conformation with enhanced preference for fibrinogen over protein C²³. Because the Na⁺ binding site involves residues near E229 (R233 and K236)²³, the effect of E229A substitution may be to lock thrombin in the Na⁺-free conformation. However, a comparison of the effects of the Na⁺ transition with the effects of the E229A mutation on the catalytic efficiency of FPA release (7-fold versus 40-fold) and on affinity for thrombomodulin (4.4-fold versus unchanged) suggests they are distinct phenomena.

* Present address: Stanford University Medical School, Division of Hematology, Stanford, California 94305-5112, USA.

FIG. 2 Anticoagulation in adult *Cynomolgus* monkeys after 10-min infusions with $2.5 \mu\text{g kg}^{-1} \text{min}^{-1}$ PCA or wild-type thrombin in 10-ml sterile isotonic saline through a peripheral vein. Blood samples were obtained by venipuncture before, during and 80 min after the infusion. All determinations were done at least in duplicate, all values are the mean of 3 or 2 separate infusions in different animals, error bars indicate standard errors. **a**, Infusion of PCA ($2.5 \mu\text{g kg}^{-1} \text{min}^{-1}$) $n=3$. Baseline aPTT was 22.4 ± 2.4 (s) and baseline fibrinogen was $215 \pm 24 \text{ mg dl}^{-1}$. **b**, Infusion of wild-type thrombin ($2.5 \mu\text{g kg}^{-1} \text{min}^{-1}$) $n=2$. Baseline aPTT was 21.6 ± 0.6 (s) and baseline fibrinogen was $249 \pm 5 \text{ mg dl}^{-1}$.

METHODS. All animals received humane care consistent with published guidelines²⁴. Activated partial thromboplastin time (aPTT) was determined using an MLA 700 automatic coagulation timer and actin FS aPTT reagent (Baxter). Plasma aPC concentration was determined by the hydrolysis of the chromogenic peptidyl substrate, S-2366 following elution of aPC captured by an EDTA-dependent monoclonal antibody immobilized on agarose beads²⁵. Concentrations of aPC were determined from a standard curve generated by dilution of pooled normal *Cynomolgus* monkey plasma and activation by Protac (American Diagnostica Inc.). Purified human aPC could not be used as a standard because monkey and human aPC were differentially recognized by the immobilized antibody. Baseline aPC levels in monkeys (1–2% of total PC) were higher than those previously reported for humans^{25,26} (0.05–0.08%) and baboons^{9,27} (0.05–0.12%), however, the precise reason for this discrepancy is uncertain. Fibrinogen levels were determined using the von Clauss method. Plasma samples (200 μl , diluted 1:10 in Owren's veronal buffer (Baxter)) or fibrinogen standards (Baxter) were incubated at 37°C for 2 min. Clotting was initiated by the addition of 100 μl Data-Fi thrombin (Baxter).



conserved among eight mammalian and avian FPA sequences (Fig. 1b). The substitution of the P5 glycine with valine in abnormal human fibrinogen Rouen results in markedly slower cleavage and a corresponding bleeding diathesis¹⁴. Despite the importance of thrombin residue E229 for recognition of fibrinogen, our results indicate E229 is dispensable for recognition of protein C by the thrombin-thrombomodulin complex. Examination of the sequences flanking the cleavage site in fibrinogen and protein C suggests that protein C may interact with thrombin in an extended conformation that is quite different from fibrinogen and independent of E229 (Fig. 1b).

To assess whether the altered substrate specificity of PCA demonstrated *in vitro* was preserved *in vivo*, PCA and wild-type thrombin were infused in Cynomolgus monkeys. Infusions of 2.5 µg kg⁻¹ min⁻¹ PCA (equivalent activated protein C (aPC) generating activity to 3 NIH U kg⁻¹ min⁻¹ wild-type thrombin) for 10 min in three monkeys caused rapid anticoagulation as assessed by a 2.3-fold prolongation of the clotting time (activated partial thromboplastin time, aPTT) (Fig. 2a). The anticoagulant effect was reversible with the aPTT returning to essentially baseline levels 80 min post-infusion. The kinetics of onset and reversal of anticoagulation were closely paralleled by the appearance and disappearance of endogenous aPC activity (with an estimated half-life of 15 min) suggesting a cause and effect relationship (Fig. 2a). About 3.5% of total plasma protein C was converted to aPC resulting in peak levels of ~4 nM aPC (assuming total plasma protein C concentration = 70 nM), levels comparable to those (6.6 nM) achieved after infusion of wild-type thrombin (2 NIH U kg⁻¹ min⁻¹, 60 min) in baboons⁹. Template bleeding time, a measure of gross platelet function, was not prolonged (Fig. 3a) indicating that the anticoagulant effect was specific and did not perturb primary haemostasis.

In contrast, infusion of wild-type thrombin also at 2.5 µg kg⁻¹ min⁻¹ (8.75 NIH U kg⁻¹ min⁻¹) for 10 min in two monkeys resulted in the generation of equivalent aPC (3.5% of total PC) and a comparable anticoagulant effect (1.8-fold prolongation of aPTT) relative to PCA despite the almost three-fold increased dose of aPC-generating activity (Fig. 2b). Consistent with previous studies in baboons⁹, the effects of thrombin's procoagulant activities were also observed after infusion of wild-type thrombin. Significant fibrinogen consumption (~20%) (Fig. 2b), marked elevation of FPA (250 nM) and factor VIII deple-

FIG. 3 Maintenance of haemostatic parameters and preservation of primary haemostasis as measured by bleeding time during anticoagulation following infusion of PCA or wild-type thrombin. Infusions were as described in Fig. 2. Samples for determination of platelet factor 4, β-thromboglobulin, platelet count, factor V and factor VIII were collected at the start of the infusion and 90 min later. Samples for determination of FPA and D-dimer were collected before infusion and after 10 min. Template bleeding times were measured before infusion and at the peak of anticoagulation (*t* = 20 min). All determinations were made at least in duplicate, all values are the mean of 3 or 2 separate infusions of different animals, error bars indicate standard errors. a, Infusion of PCA (2.5 µg kg⁻¹ min⁻¹) *n* = 3. b, Infusion of wild-type thrombin (2.5 µg kg⁻¹ min⁻¹) *n* = 2.

METHODS. Platelet factor 4, β-thromboglobulin, FPA and D-dimer levels were determined using commercial ELISA kits (Diagnostica Stago). Factor V and factor VIII levels were determined by clotting assays based on the prothrombin time and aPTT⁹. Sample plasma diluted 1:10 in TBS containing 0.1% BSA (50 µl) was mixed with plasma deficient in factor V or factor VIII (George King Biomedical) (50 µl) and thromboplastin (100 µl) or actin FS aPTT reagent (Sigma) (100 µl), respectively, and incubated at 37 °C for 5 min. Clotting was initiated by the addition of 100 µl of 20 mM CaCl₂. Levels of FV and FVIII are expressed as a percentage of the values obtained for pooled normal plasma from Cynomolgus monkeys. A blood pressure cuff was inflated to 40 mm Hg and the bleeding time was measured on the volar surface of the forearm using a commercial device (Surgicutt).

TABLE 1 Specificity of PCA compared to wild-type thrombin

Assay	WT*	PCA* (E229A)	Relative specificity
Fibrinogen Aα chain†			
<i>k</i> _{cat} (s ⁻¹)	57 ± 1	4.53 ± 0.05	
<i>K</i> _m (µM)	3.55 ± 0.01	11.3 ± 0.1	
<i>k</i> _{cat} / <i>K</i> _m (s ⁻¹ µM ⁻¹)	16.1 ± 0.4	0.40 ± 0.01	2.5%
Clotting activity (human plasma)‡			
[IIa] _{for t = 60 s} (nM)	3.9 ± 0.7	58 ± 13	6.7%
S-2238 (H-D-Phe-Pip-Arg-pNA)§			
<i>k</i> _{cat} (s ⁻¹)	51 ± 2	38 ± 1	
<i>K</i> _m (µM)	2.95 ± 0.46	50.2 ± 2.7	
<i>k</i> _{cat} / <i>K</i> _m (s ⁻¹ µM ⁻¹)	17.3 ± 3.4	0.76 ± 0.06	4.4%
Platelet aggregation 			
EC ₅₀ (nM)	0.4 ± 0.1	31 ± 8	1.3%
Platelet thrombin receptor peptide*			
<i>k</i> _{cat} (s ⁻¹)	34 ± 5	25 ± 2	
<i>K</i> _m (µM)	35.5 ± 9.2	181 ± 19	
<i>k</i> _{cat} / <i>K</i> _m (s ⁻¹ µM ⁻¹)	0.96 ± 0.39	0.14 ± 0.03	14.6%
Thrombomodulin affinity*			
<i>K</i> _D (memb TM) (nM)	4.3 ± 0.6	3.8 ± 1	88%
<i>K</i> _D (soluble TM) (nM)	18 ± 10	29 ± 20	161%
Protein C‡			
<i>k</i> _{cat} (s ⁻¹)	0.65 ± 0.06	0.19 ± 0.01	
<i>K</i> _m (µM)	3.2 ± 0.5	2.00 ± 0.13	
<i>k</i> _{cat} / <i>K</i> _m (s ⁻¹ µM ⁻¹)	0.20 ± 0.05	0.095 ± 0.009	47.5%
Antithrombin III inhibition**			
<i>k</i> _{on} (s ⁻¹ M ⁻¹)	5.42 ± 0.30 × 10 ³	7.83 ± 0.67 × 10 ²	14.4%
Plasma Stability in vitro††			
<i>t</i> _{1/2} (s)	30	180	16.6%

Comparison of the specificity of wild-type and E229A thrombin (PCA) in a variety of *in vitro* assays with standard errors.

* The human prothrombin coding sequence from cDNA clone BS(KS)hFII (provided by R. MacGillivray, University of British Columbia) was expressed using pRC/CMV (Invitrogen Corp.)¹². The E229A mutation was constructed by oligonucleotide-directed mutagenesis. Prothrombin was secreted from stably expressing BHK-21 cell lines grown in serum-free DMEM in expanded-surface roller bottles. Conditioned medium was filtered and concentrated using a tangential flow filtration system (Pellicon, Millipore). Prothrombin was purified by anion exchange using DEAE-sepharose¹⁵ and processed to thrombin by *Echis carinatus* venom. Thrombin was purified by cation exchange using Amberlite CG50¹⁶ with an overall yield of 28%. Purity was assessed to be greater than 90% by silver-stained SDS-PAGE. Protein concentration was determined by BCA protein assay (Pierce) and by a thrombin ELISA using PPACK to capture thrombin on the microtitre plate surface¹⁷ and polyclonal antibodies (Dako) to detect thrombin.

† Kinetic analysis of FPA release was determined after separation of FPA by reverse-phase HPLC¹⁸ and quantification by comparison with purified FPA (Sigma). Wild-type thrombin concentration was 0.3 nM, PCA was 7 nM. Kinetic parameters were determined by curve fitting to differential progress curves¹⁹ generated at two fibrinogen concentrations, 1.4 and 21 µM.

‡ Human plasma (150 µl) and 10–200 nM WT or PCA (150 µl in 2 × TBS plus 0.1% BSA) were prewarmed at 37 °C for 1 min in a manual fibrometer (Becton-Dickinson), then mixed. Concentrations required for a clotting time of 60 s were determined from plots of concentration versus clotting time.

§ Kinetic analysis of the hydrolysis of the chromogenic FPA mimetic S-2238 (H-D-Phe-Pip-Arg-pNA) (0–100 µM) was done using 0.25 nM WT and 1.0 nM PCA. Kinetic parameters were calculated by fitting to the Michaelis-Menten equation by nonlinear regression.

|| Platelet aggregation induced by PCA (6.25–250 nM) or WT (2.5–50 nM) was determined using citrated human platelet-rich plasma in a dual channel aggregometer (Chrono-Log Corp.). Extent of aggregation was determined from the area under the tracing 0–1.5 min after the addition of thrombin. The concentrations required for half-maximal stimulation of aggregation (EC₅₀) were determined from plots of the concentration dependence of aggregation.

* Kinetic analysis of the hydrolysis of a peptidyl substrate derived from the platelet thrombin receptor³ was determined after separation of the products by reverse-phase HPLC. The substrate, Ac-Ala-Thr-Leu-Asp-Pro-Arg-Ser-Phe-Leu-Leu-Arg-Asn-Pro-Asn-Asp-Lys-Tyr-Glu-Pro-Phe-Trp-NH₂ was synthesized by Peninsula Laboratories Inc. Kinetic parameters were calculated by curve fitting to the Lineweaver-Burk transformation of the Michaelis-Menten equation.

* Affinity of PCA or WT (0.1–40 nM) for soluble recombinant thrombomodulin (solubleTM, 0.5 nM) (American Diagnostica Inc.) or full-length thrombomodulin in a lysate prepared from TMnc cells expressing recombinant thrombomodulin²⁰ (membTM, 0.5 nM) was determined from the concentration dependence of protein C (1.0 µM) activation²¹. Protein C activation was monitored by hydrolysis of the chromogenic peptidyl substrate S-2366 (200 µM) and quantified by comparison to a standard curve generated with purified aPC (Haematologic Technologies).

‡ Kinetic analysis of protein C activation by PCA or WT (20 nM) was determined in the presence of full-length thrombomodulin (0.5 nM). Protein C concentration was varied between 0.5–1.0 µM. Kinetic parameters were calculated by curve fitting to the Michaelis-Menten equation.

** Second-order rate constants of antithrombin III inhibition were determined as described²² using purified ATIII (Enzyme Research Laboratories).

†† Plasma half-life was determined in fibrinogen-deficient plasma (George King Biomedical) by monitoring S-2238 hydrolysis versus time.

tion (66%) (Fig. 3b) indicated activation of the coagulation cascade and elevation of β -thromboglobulin and platelet factor 4 levels (Fig. 3b) indicated substantial platelet activation had occurred. Notably, apart from a modest increase in FPA levels (40 nM), no evidence of thrombin's procoagulant activities was observed after PCA infusion (Figs 2a and 3a), even at a high dose ($4.5 \mu\text{g kg}^{-1} \text{min}^{-1}$; 5.5-fold prolongation of aPTT, data not shown), suggesting a broad therapeutic range if PCA was to be used as an anticoagulant.

Our results indicate that thrombin, a key regulator of blood coagulation with opposing procoagulant and anticoagulant activities, can be engineered to function *in vivo* as a potent anticoagulant that may possess a superior therapeutic profile with reduced potential for bleeding complications. □

Received 31 July; accepted 13 September 1995.

1. Jackson, C. M. & Nemerson, Y. A. *Rev. Biochem.* **49**, 765–811 (1980).
2. Mann, K. G. & Lundblad, R. L. in *Hemostasis and Thrombosis* (eds Colman, R. W., Hirsh, J., Marder, V. J. & Salzman, E. W.) 148–161 (Lippincott, Philadelphia, PA, 1987).
3. Vu, T. K. H., Wheaton, V. I., Hung, D. T., Charo, I. & Coughlin, S. R. *Nature* **353**, 674 (1991).
4. Lorand, L. & Radek, J. T. in *Thrombin Structure and Function* (ed. Berliner, L. J.) 257–270 (Plenum, Columbus, OH, 1992).
5. Gailani, D. & Broze, G. J. Jr *Science* **253**, 909–912 (1991).

6. Mann, K. G., Jenny, R. J. & Krishnaswamy, S. A. *Rev. Biochem.* **57**, 915–956 (1988).
7. Esmon, C. T. *Science* **235**, 1348–1352 (1987).
8. Comp, P. C., Jacobs, R. M., Ferrell, G. L. & Esmon, C. T. *J. clin. Invest.* **70**, 127–134 (1982).
9. Hanson, S. R. et al. *J. clin. Invest.* **92**, 2003–2012 (1993).
10. Wu, Q. et al. *Proc. natn. Acad. Sci. U.S.A.* **88**, 6675–6779 (1991).
11. Tsiang, M. et al. *J. biol. Chem.* **270**, 16854–16863 (1995).
12. Stubbs, M. T. et al. *Eur. J. Biochem.* **206**, 187–195 (1992).
13. Banfield, D. K. & MacGillivray, R. T. A. *Proc. natn. Acad. Sci. U.S.A.* **89**, 2779–2783 (1992).
14. Ni, F., Konishi, Y., Bullock, L. D., Rivetna, M. N. & Scheraga, H. A. *Biochemistry* **28**, 3106–3119 (1989).
15. Borowski, M., Furie, B. C., Goldsmith, G. H. & Furie, B. *J. biol. Chem.* **260**, 9258–9264 (1985).
16. Fenton, J. W., Fasco, M. J. & Stackrow, A. B. *J. biol. Chem.* **252**, 3587–3598 (1977).
17. Tsiang, M. et al. *J. biol. Chem.* **270**, 19370–19376 (1995).
18. Higgins, D. L. *J. biol. Chem.* **258**, 9276–9282 (1983).
19. Graycar, T. P. & Estell, D. A. *J. cell. Biochem.* **11c** (suppl.), 234 (1987).
20. Tsiang, M., Lentz, S. R. & Sadler, J. E. *J. biol. Chem.* **267**, 6164–6170 (1992).
21. Owen, W. G. & Esmon, C. T. *J. biol. Chem.* **256**, 5532–5535 (1981).
22. Ye, J. et al. *J. biol. Chem.* **269**, 17965–17970 (1994).
23. Dang, Q. D. et al. *Proc. natn. Acad. Sci. U.S.A.* **92**, 5977–5981 (1995).
24. National Academy of Sciences *Guide for Care and Use of Laboratory Animals* **85-23** (National Institutes of Health, Washington DC, 1985).
25. Orthner, C. L., Kolen, B. & Drohan, W. N. *Thromb. Haemostasis* **69**, 441–447 (1993).
26. Gruber, A. & Griffin, J. H. *Blood* **79**, 2340–2348 (1992).
27. Gruber, A. et al. *Circulation* **84**, 2454–2462 (1991).

ACKNOWLEDGEMENTS. We thank P. Mack, DVM (Tulane Regional Primate Research Center) for assistance with animal studies, C. L. Orthner (American National Red Cross) for the gift of the monoclonal antibody against protein C and J. Toole, R. Wagner and M. Matteucci for helpful comments regarding this manuscript.

A protein catalytic framework with an N-terminal nucleophile is capable of self-activation

James A. Brannigan*, Guy Dodson*†, Helen J. Duggleby*, Peter C. E. Moody*‡, Janet L. Smith§, Diana R. Tomchick§ & Alexey G. Murzin||

* Chemistry Department, University of York, York YO1 5DD, UK

† NIMR, The Ridgeway, London NW7 1AA, UK

‡ Present address: Biochemistry Department, University of Leicester, Leicester LE1 7RH, UK

§ Department of Biological Sciences, Purdue University, West Lafayette, Indiana 47907, USA

|| MRC Centre for Protein Engineering, Hills Road, Cambridge, CB2 2QH, UK

THE crystal structures of three amidohydrolases have been determined recently^{1–3}: glutamine PRPP amidotransferase (GAT), penicillin acylase, and the proteasome. These enzymes use the side chain of the amino-terminal residue, incorporated in a β -sheet, as the nucleophile in the catalytic attack at the carbonyl carbon. The nucleophile is cysteine in GAT, serine in penicillin acylase, and threonine in the proteasome. Here we show that all three enzymes share an unusual fold in which the nucleophile and other catalytic groups occupy equivalent sites. This fold provides both the capacity for nucleophilic attack and the possibility of autocatalytic processing. We suggest the name Ntn (N-terminal nucleophile) hydrolases for this structural superfamily of enzymes which appear to be evolutionarily related but which have diverged beyond any recognizable sequence similarity.

The fold common to all three enzymes, illustrated schematically in Fig. 1, is a four-layer $\alpha + \beta$ structure with two antiparallel β -sheets, one essentially flat (I), the other slightly twisted (II). The β -sheets are packed in a unique arrangement in which one is rotated relative to the other through a positive dihedral angle of $+30^\circ$, in contrast to the usual value of -30° strongly favoured by the typical β -sheet twist⁴ (Fig. 2). Unlike the twisted β -sheet, the flat sheet surface has cavities that can accommodate the side-chains from the other sheet. This intercalation, which brings the β -sheets close together, is facilitated by an unusually large proportion of small and non-branched residues.

In the proteasome, the common fold covers nearly all of each subunit (Fig. 1). In penicillin acylase (PA), the common fold is embedded in its B-subunit but is interrupted by various excursions, some of which are very large. One of the β -sheets (II) is, for example, greatly extended to ten strands in PA. In GAT, the common fold occupies the N-terminal domain, which contains the glutamine amidotransferase activity. The topology of this domain is slightly different from those of PA and the proteasome (PRO) (Fig. 1), although superposition of the PA and GAT structures shows that they match well (Fig. 3).

Structural alignment of the active sites of Ntn hydrolases shows that elements of their catalytic centres are equivalent (Figs 3 and 4). Catalytic roles have been proposed for specific chemical groups in each of the enzymes^{2,3,5}. The nucleophile, proton donor, one element of the oxyanion hole, and the substrate-binding pocket occupy topologically equivalent sites. An overlap based on catalytically relevant atoms gives an r.m.s. deviation of 0.7 Å for superposition of SerB1 O_γ, SerB1 N, AlaB69 N and AsnB241 N_{δ2} of PA with Cys1 Sy, Cys1 N, Gly103 N and Asn102 N_{δ2} of GAT. The common catalytic framework of PA, PRO and GAT thus suggests that the three enzymes share a common catalytic mechanism, and that the reactions proceed with the same chirality in the intermediates, even though the nucleophiles are different. The structural variation and sequence diversity in these enzymes have been seen before in other superfamilies, such as the α/β hydrolases^{6,7} and the restriction endonucleases⁸.

In all of the Ntn hydrolases, post-translational processing from inactive precursors generates, in the residue with the nucleophilic side chain, a free α -amino group which evidently participates in the catalytic process. The β -subunits of PRO and some GATs are activated by the removal of a short prosequence. The active PA is a heterodimer of A- and B-subunits, derived from a single-chain precursor. In the case of PRO, this process is autocatalytic⁹. For GAT and PA, the possibility of autocatalysis has been suggested from observations that expression of precursors in a different host results in active enzyme and that the processing can be affected by mutations far from the cleavage sites^{10–14}. This autoproteolysis might help to ensure that the enzyme is correctly folded or active only at its correct cellular location.

The framework of Ntn hydrolases suggests a probable mechanism for autocatalysis. The precursor residue adjacent to the nucleophile might bind in the active site to fit its side chain in the substrate-binding pocket and carbonyl oxygen in the oxyanion hole. This would simultaneously induce conforma-

tional strain at the cleaving peptide and allow attack by the nucleophile group at the carbonyl carbon. The resulting tetrahedral intermediate would collapse into the acyl-(thio)ester intermediate, which would then be hydrolysed. An analogous case of autocatalytic cleavage by serine, in which conformational strain generated by an unusual β -sheet framework appears to be a factor, is seen in the pyruvoyl-dependent histidine decarboxylase¹⁵.

There may be other Ntn hydrolases. An N-terminal threonine has been identified as the catalytic residue of heterodimeric lyso-

somal glycosylasparaginase (aspartylglycosylaminase; AGA)¹⁶. The single-chain precursors of this and related enzymes undergo autocatalytic cleavage just before this invariant threonine^{17,19}. A similar maturation event is evident in γ -glutamyl transferase (GGT)^{20,21}. The catalytic nucleophile of GGT is yet to be identified²², and we predict it to be the invariant threonine at the N terminus of the B-chain. If GGT is an Ntn hydrolase, it will provide biochemical links to the others: GGT, like GAT, can use glutamine as substrate²⁰. Some members of the GGT and PA families act as cephalosporin acylases²³. GGT has a

FIG. 1 Topological diagrams for the Ntn hydrolases. *a*, The notation scheme for the α -helices (circles) and β -strands (triangles) that are spatially equivalent in the proteasome (PRO) subunits, penicillin acylase (PA) and glutamine PRPP amidotransferase (GAT). Secondary structural elements with common connectivity in the three proteins are shown in black and are connected. Structural elements that are spatially equivalent but have different connectivity are grey. *b*, The PRO subunits: the catalytic β -subunit; the catalytically inactive α -subunit has an additional α -helix at the N terminus which mediates proteasome assembly⁹. *c*, The B-chain of PA. This has several excursions from the common fold, two of which form separate domains, D. *d*, The N-terminal domain of GAT, which has a slightly different topology from those of PRO and PA. Diagrams of the individual proteins include all structural elements constituting each domain. Those that are not spatially equivalent to the common core shown in *a*, are depicted as open symbols. The structural alignment of PRO with PA and GAT was derived from the stereoscopic figures in ref. 3. Topologically equivalent residues in the β -sheet are: β 1, 1–7 (PRO- β), 1–7 (PA-B), 1–7 (GAT); β 2, 12–18, 16–22, 66–72; β 3, 35–37, 58–60, 87–89; β 4, 41–47, 63–69, 97–103; β 5, 97–102, 177–182, 158–163; β 6, 110–115, 188–193, 168–173; β 7, 123–127, 235–239, 189–193. Other spatially equivalent β -sheet residues are: β A, 173–178, 483–488, 31–36; β B, 184–188, 497–501, 41–45; β C, 188–121, 227–230, 213–216. There are no preserved residues in the four sequences, including the proteasome α -subunit.

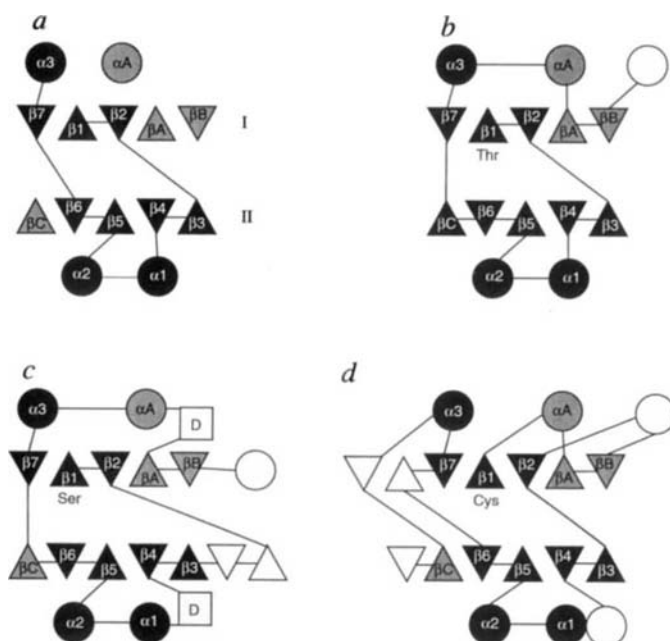


FIG. 2 The front (*a*, *b*) and side (*c*, *d*) views of the secondary structure packing in the Ntn hydrolase fold. Only the common fold regions and the N-terminal residue of PA (*a*, *c*) and GAT (*b*, *d*) are shown. Note the unusual positive dihedral angle between the direction of strands in the top β -sheet I, which is essentially flat, and the bottom β -sheet II, which is slightly twisted. The same packing is evident in the proteasome subunits (for comparison see Fig. 2 in ref. 3).

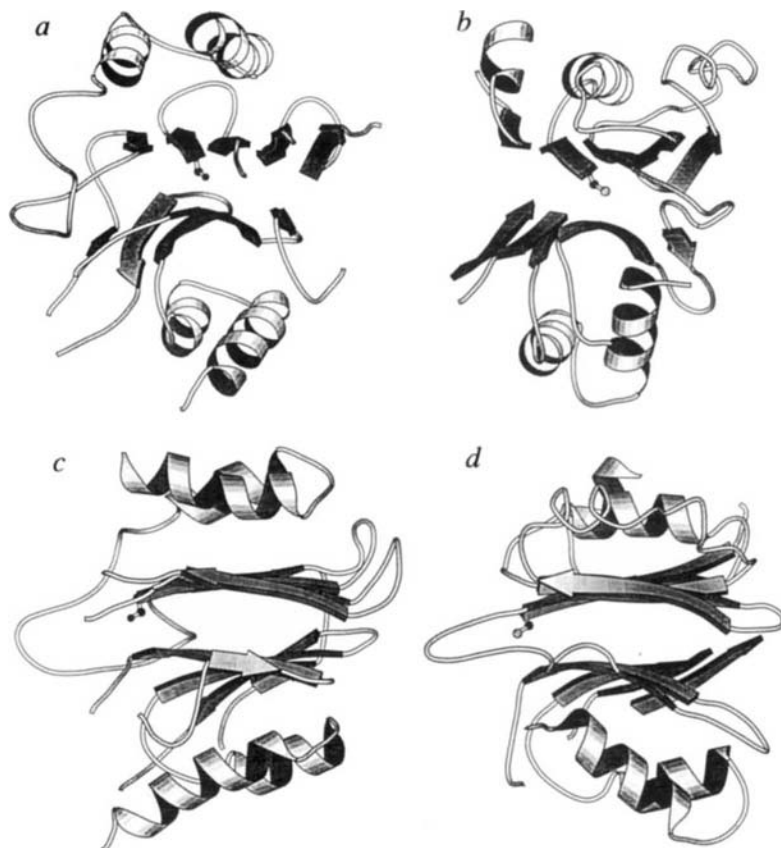


FIG. 3 Superposition of the secondary structure packing in PA (thick line) and GAT (thin line) illustrated in a stereoscopic view. The structural elements shown are those seen in Fig. 1a. The position of key atoms of the catalytic centre are marked with larger spheres (GAT Cys1 S γ and N, Gly103 N; PA SerB1 O γ and N, AlaB69 N) and are shown in greater detail in Fig. 4. The common core of 113 C α atoms (76 in topologically equivalent and 37 in other structurally analogous regions) has an r.m.s. deviation of 2.0 Å; these do not include the atoms of helix α 1, which appears in a similar location in both structures but does not overlap well.

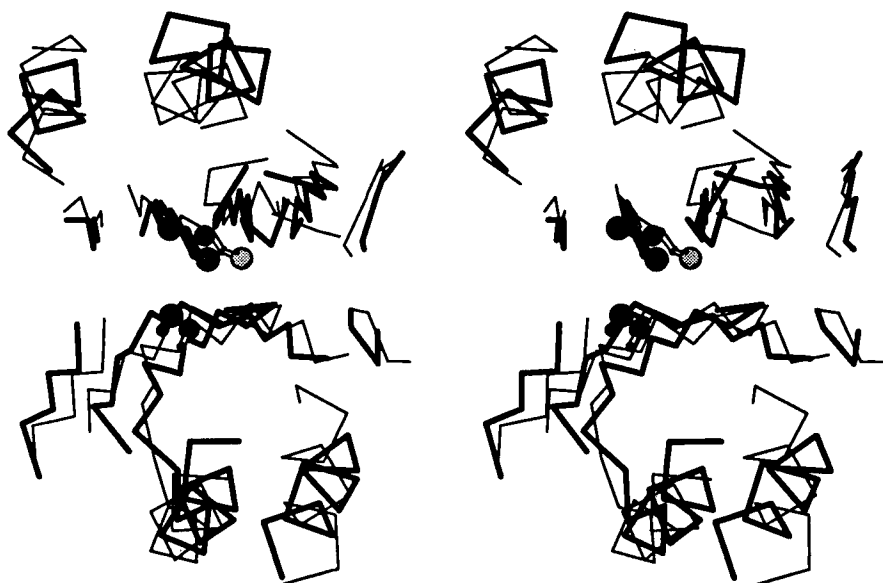
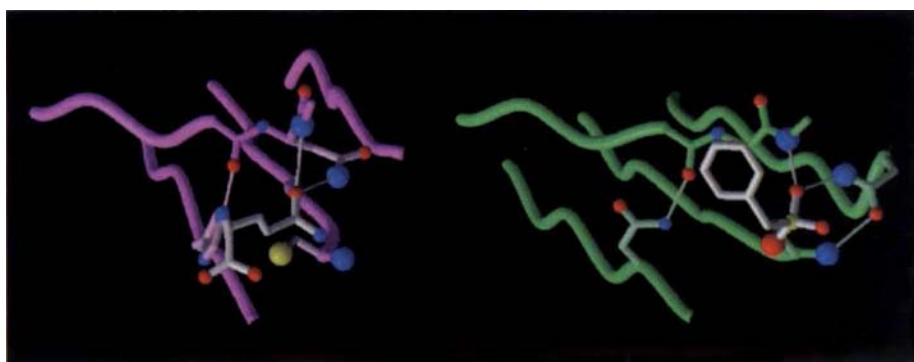


FIG. 4 Architecture of the active site of Ntn hydrolases. The active site regions of PA (green) and GAT (magenta) are shown, including the side chains for SerB1, AsnB20 and AsnB241 of PA, and Cys1, His70 and Asn102 of GAT. Main chain atoms of PA SerB67, AlaB69 and GAT Asn102, Gly103 are also shown. Hydrogen bonds are drawn as thin white lines; nitrogen atoms are blue, oxygen red and sulphur yellow. For details of the bound substrates, see below. The nucleophile, a thiol or hydroxyl group, and the proton donor, the α -amino group, are both in the N-terminal residue of strand β 1 in the top sheet of this view. The first element of the oxyanion hole is the backbone NH group at the end of strand β 4 in the bottom sheet (GAT, Gly103; PA, AlaB69; in PRO the structurally equivalent Gly47 plays the same role³). This NH is close enough to act because the β -sheets are packed very close, and is available for hydrogen bonding to the substrate because the adjacent strand, β 3, is shorter than β 4. Strand β 2 helps to fix this architecture by making hydrogen-bonding contacts to both β 1 and to β 4 near the oxyanion hole. There is a conserved hydrogen bond from a side chain in β 2 (PA, AsnB20; GAT, His70; and, by analogy, PRO, Thr16) to the main chain C=O in β 4 (PA, SerB67; GAT, His101; PRO, Ile45); this is shown for PA and GAT. In PA, a second contributor to the oxyanion hole is donated by the side chain of AsnB241, whereas in GAT the analogous group of Asn102 comes from a different region of the structure. In PRO, the conserved Ser129 appears to be structurally equivalent to AsnB241 in PA, and its hydroxyl group is likely to play the same role. Mutation of this residue to alanine



significantly reduces catalytic activity and post-translational processing²⁸. The gap left by the shorter strand, β 3, forms part of the substrate-binding pocket. This site in GAT is surrounded by residues invariant in class II glutamine amidotransferases²⁹ and is proposed as the binding site⁵ for glutamine (white bonds). It is modelled into the active site with C δ positioned for nucleophilic attack by the thiolate of Cys1, O ϵ in the oxyanion hole, N ϵ within hydrogen-bonding distance of the α -amino group of Cys1 and the charged amino and carboxyl groups forming salt bridges with invariant Asp and Arg side chains (not shown⁵). The aromatic ring of the PA inhibitor phenylmethylsulphonate (shown here in white, covalently attached through the serine), and the norLeu side chain at the P $_1$ site of the PRO inhibitor Ac-Leu-Leu-norLeu, also appear in this location^{2,3}.

single-chain precursor, like PA and AGA, and an N-terminal threonine in the mature enzyme, like AGA and PRO.

Finally, an Ntn hydrolase domain might reside within the intein, a self-excision element in protein splicing^{24,25}. All known inteins have an N-terminal Cys or Ser residue, and the C-terminal exteins have an N-terminal Cys, Ser or Thr residue. Both nucleophile residues are essential for splicing. In accordance with experimental observation^{26,27}, the catalytic properties of the Ntn hydrolases suggest a mechanism for the complex process of protein splicing. Thus the intein (with the Ntn structure) autocatalytically cleaves the upstream junction, activating the Ntn so it can cleave the downstream peptide junction, either directly, or by activating the side chain amido group of the invariant C-terminal asparagine of the intein. Between these events, the second nucleophile forms the observed branched, acylated intermediate with the N-terminal extein. This hypothetical mechanism takes advantage of the fact that the Ntn hydrolase fold

can simultaneously bring about autocatalysis and generate an active nucleophile.

Note added in proof: The crystal structure of AGA, predicted here to be a Ntn hydrolase, has been determined³⁰ and reveals essentially the same catalytic framework (J. Rouvinen, personal communication). The crystal structure determination of GGT is also in progress³¹. □

Received 7 July; accepted 6 October 1995.

- Smith, J. L. et al. *Science* **264**, 1427–1433 (1994).
- Duggleby, H. J. et al. *Nature* **373**, 264–268 (1995).
- Löwe, J. et al. *Science* **268**, 533–539 (1995).
- Chothia, C. & Janin, J. *Proc. natn. Acad. Sci. U.S.A.* **78**, 4146–4150 (1981).
- Smith, J. L. *Biochem. Soc. Trans.* **23**, 894–898 (1995).
- Dodson, G. G., Lawson, D. M. & Winkler, F. K. *Faraday Discuss.* **93**, 95–105 (1992).
- Ollis, D. L. et al. *Protein Engng.* **5**, 197–211 (1992).
- Aggarwal, A. K. *Curr. Opin. struct. Biol.* **5**, 11–19 (1995).
- Zwickl, P. et al. *Nature struct. Biol.* **1**, 765–769 (1994).
- Sizmann, D. et al. *Eur. J. Biochem.* **192**, 143–151 (1990).

11. Mäntsälä, P. & Zalkin, H. *J. biol. Chem.* **259**, 14230–14236 (1984).
12. Souciet, J.-L. et al. *J. biol. Chem.* **263**, 3323–3327 (1988).
13. Zhou, G. et al. *J. biol. Chem.* **267**, 7936–7942 (1992).
14. Bruns, W. et al. *J. molec. Appl. Genet.* **3**, 36–44 (1985).
15. Gallagher, T. et al. *J. molec. Biol.* **230**, 516–528 (1993).
16. Kaartinen, V. et al. *J. biol. Chem.* **266**, 5860–5869 (1991).
17. Mononen, I. et al. *FASEB J.* **7**, 1247–1256 (1993).
18. Lough, T. J. et al. *Pl. molec. Biol.* **19**, 391–399 (1992).
19. Tarentino, A. L. et al. *Archs. Biochem. Biophys.* **316**, 399–406 (1995).
20. Tate, S. S. & Meister, A. *Meth. Enzym.* **113**, 400–419 (1985).
21. Suzuki, H. et al. *J. Bact.* **171**, 5169–5172 (1989).
22. Smith, T. K. et al. *Proc. natn. Acad. Sci. U.S.A.* **92**, 2360–2364 (1995).
23. Matsuda, A. et al. *J. Bact.* **169**, 5821–5826 (1987).
24. Wallace, C. J. A. *Protein Sci.* **2**, 697–705 (1993).
25. Colston, M. J. & Davis, E. O. *Molec. Microbiol.* **12**, 359–363 (1994).
26. Xu, M.-Q. et al. *Cell* **75**, 1371–1377 (1993).
27. Xu, M.-Q. et al. *EMBO J.* **13**, 5517–5522 (1994).
28. Seemüller, E. et al. *FEBS Lett.* **359**, 173–178 (1995).
29. Zalkin, H. *Adv. Enzym.* **66**, 203–309 (1993).
30. Oinonen, C. et al. *Nature struct. Biol.* (in the press).
31. Kumagai, H. et al. *J. molec. Biol.* **234**, 1259–1262 (1993).

ACKNOWLEDGEMENTS. We thank D. Edwards for technical assistance and C. Chothia, D. H. Jones and J. Colston for discussion and ideas. This work was supported by SERC (Protein Engineering), the MRC, the NIH, and the Lucille P. Markey Foundation.

CORRECTIONS

MIF is a pituitary-derived cytokine that potentiates lethal endotoxaemia

J. Bernhagen, T. Calandra, R. A. Mitchell, S. B. Martin, K. J. Tracey, W. Voelker, K. R. Manogue, A. Cerami & R. Bucala

Nature **365**, 756–759 (1995)

In Fig. 2 legend, the orientation of the downstream primers was inadvertently reversed: the sequences are correct but they run from 5' to 3' and not from 3' to 5' as published. □

Crucial role of the pre-T-cell receptor α gene in development of $\alpha\beta$ but not $\gamma\delta$ T cells

Hans Jörg Fehling, Anna Krotkova, Claude Saint-Ruf & Harald von Boehmer

Nature **375**, 795–798 (1995)

In Table 1 of this Letter, the absolute numbers of thymocytes listed for phenotypes $CD4^+8^+3^{low}25^+$, $CD4^+8^+\delta^+$ and $CD4^+8^+\delta^+$ were ten times lower than the published values. The reported numbers for the other phenotypes are correct. Also, in the last sentence on page 797, the text should read "... in the presence of TCR-positive cells, the pre-TCR is required for the transition of $CD4^+CD8^+25^+$ $\alpha\beta$ T-cell precursors ..." (instead of $CD4^+CD8^+25^+$ as published). □

Immunological function of a defined T-cell population tolerized to low-affinity self antigens

Kazuhiro Kawai & Pamela S. Ohashi

Nature **374**, 68–69 (1995)

SINCE the publication of this Letter, it has come to our attention that a model proposed by Grossman and Paul (*Proc. natn. Acad. Sci. U.S.A.* **89**, 10365–10369; 1992) is similar to our interpretation of these experiments. Our data provide experimental evidence for this hypothesis. □

Regulation of neural induction by the Chd and Bmp-4 antagonistic patterning signals in *Xenopus*

Yoshiki Sasai, Bin Lu, Herbert Steinbeisser & Eddy M. De Robertis

Nature **376**, 333–336 (1995)

THE PCR primer sequences for NF-M and NCAM in the legend of Fig. 1 were accidentally omitted. The primer sequences for NF-M are: F, 5'-GAACAGTACGCCAAGCTGACTG, and R, 5'-GCAGCAATTTCTATATCCAGAG (28 cycles); those for NCAM (accession no. M25696) are: F, 5'-GCGGGTACCTTCTAATAGTCAC, and R, 5'-GGCTTGGCTGTGGTTCTGAAGG (28 cycles). We thank R. Harland for pointing out our mistake. □

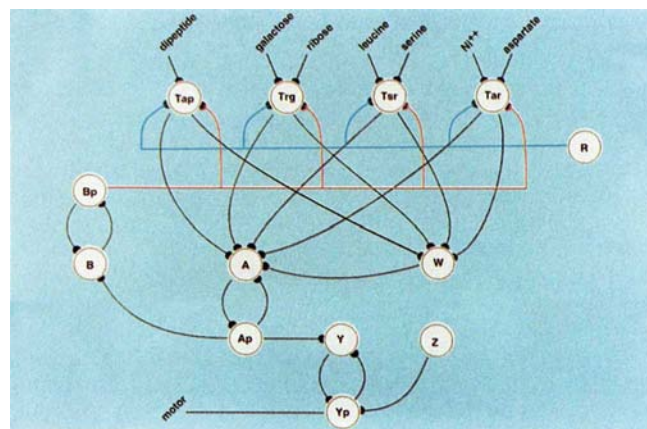
ERRATA

Protein molecules as computational elements in living cells

Dennis Bray

Nature **376**, 307–312 (1995)

A FAULT in the reproduction of Fig. 4 of this Review Article appeared in the UK edition of *Nature*. The correct figure is shown here. □



A base pair between tRNA and 23S rRNA in the peptidyl transferase centre of the ribosome

Raymond R. Samaha, Rachel Green & Harry F. Noller

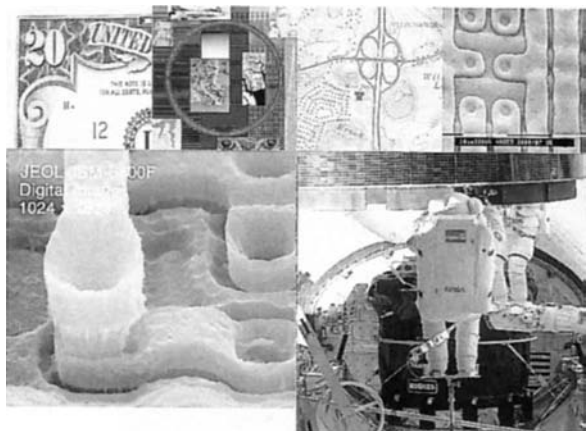
Nature **377**, 309–314 (1995)

FIGURE 6a of this Article contained a typographical error: the number in the first column of the third row should be 0.03. □

Cybernation

This edition covers new computer hardware and software, including a portable workstation, data-logging and acquisition hardware, imaging software, and on-line and CD-ROM information.

ALDEN Electronics has released a combination GPIB (general purpose interface bus)/SCSI (small computer system interface) interface for the Alden 9315CTP printer, which produces high-resolution, **photographic-quality, black and white images** that have up to 256 shades of grey per pixel (*Reader Service No. 100*). The printer uses direct thermal printing technology and generates photographic quality images of up to 25.6 cm wide and up to 38 m in length. The image is said to be pro-



Alden's GPIB/SCSI printer uses thermal printing technology for high-quality, black and white images.

duced quickly and cleanly — there is no developing, fixing, wet chemistry or waste disposal. It has only three moving parts, which allows it to operate in offices, moving vehicles and hostile field environments — no ink cartridges, ribbons or toner are required. In addition to a GPIB/SCSI Interface, the printer is also available with parallel and SEM (scanning electron microscope) slow-scan interfaces and an optional SCSI/parallel, dual interface. It comes with application software to print images produced in TIFF format. It can also function as a text-only printer, and has a high-speed black and white mode for such items as line art or images from computer-aided design systems where tone shade reproduction is not required. A choice of three different imaging materials is offered: a white plastic-based material for photographic quality images, a white fibre based material for lower cost or draft-quality images and a transparent media (PolaTherm from Polaroid) for overhead transparencies.

Univision Technologies has introduced the combination of Falcon's VGA **frame grabbers** and Media Cybernetics' Image-Pro Plus 1.3 for Windows, which is

designed to provide an integrated solution to complex image-processing problems (*Reader Service No. 101*). The program utilizes a single-slot VL or PCI bus frame grabber with built-in VGA compatibility and optional on-board DSP processing. The system offers eight-bit, 256 grey level per pixel acquisition into an inexpensive VGA display. Distinguishing features include eight-bit, independent, VGA-overlay planes, in colour or grey-scale, and an additional eight-bit planes of VGA image-storage VRAM memory. It supports nondestructive graphics or text overlays at resolutions from 640×480 up to $1,280 \times 1,024$. Moreover, acquired 30-Hz RS-170 and CCIR images are displayed on a single VGA monitor at flicker-free, non-interlaced refresh rates of up to 72 Hz. The program allows easy capture, enhancement, analysis and output of eight-bit images, and data can be exported to a spreadsheet program or printer. The program is capable of counting and sizing, analy-

sis, enhancement, densitometry, calibration, annotation, measurement, two-dimensional FFT, pseudo-colouring and contrast enhancement. The frame grabber, the software and a video camera can convert a standard VL or PCI local-bus PC into an image-processing system.

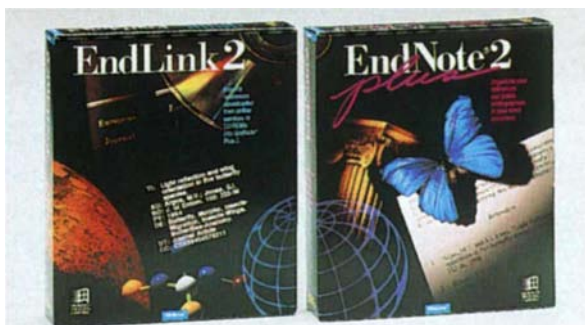
Alias/Wavefront has joined Silicon Graphics to introduce Indigo² Impact, a new line of **three-dimensional graphics and imaging workstations** (*Reader Service No. 102*). The two- and three-dimensional family of products has been optimized to take advantage of the new workstations. Alias/Wavefront products running on the system can now work in an interactive manner with immediate feedback and the ability to make adjustments in real-time. Software options for molecular visualization and modelling capabilities, include Biosym, Molecular Simulations and Tripos. The workstation is based on the MIPS R8000



Molecular modelling and visualization.

processor, which is said to offer a combination of supercomputing power and advanced graphics, and should prove useful for high-performance texture mapping. The workstation is designed to allow for real-time analysis of molecular properties, including property mapping, information filtering, property clipping and colour-coded screening of molecular properties on solid surfaces. To investigate the structural origin of a molecular function, transparency may be employed to remove portions of the molecular surface. The window that displays stereo images may be selected, while the rest of the screen remains non-stereo for easier use of other applications, avoiding blurred menu buttons and text editors.

Niles and Associates offers EndNote Plus 2.0, a **bibliographic utility** designed for use by researchers, students and academic writers. It is a collection of about 200 predefined bibliographic styles in more than 15 different disciplines (*Reader Service No. 103*). The new term-list feature in EndNote Plus 2.0 helps users keep track of important terms for data entry and retrieval. Users can create up to 31 term lists per library (such as for authors, keywords or other fields), and each list can hold thousands of names or



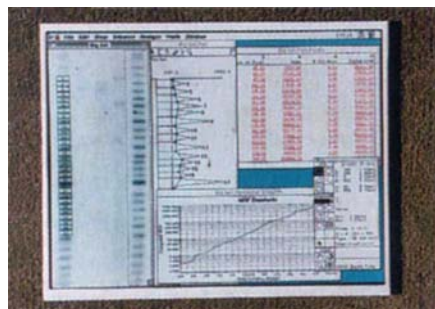
EndLink 2.0 and End Note 2.0 from Niles and Associates.

terms. Term lists can be linked to different fields and when the user is typing in a particular field, he or she can instantly call up the list linked to that field. The global editing feature helps researchers keep an organized database. Users can add text to all or just some references automatically. (Text can be appended to the beginning or end of a field, or can be set to replace the entire contents of a field.) A find-and-replace function allows the user to fix errors or standardize terms. The user can modify any of the filters, or create their own filters, specifying exactly what data they want imported, and into which fields.

Tadpole Technology has launched the SPARCbook 3 range of **workstation-class portable computers** (*Reader Service No. 104*). Although the size of a laptop computer and weighing less than seven pounds, the products include many features designed to ease mobile computing. This computer is aimed at scientific users requiring the power of their workstations while working away from their desks. It is said to offer the same system performance and features of a fully functional workstation, including ISDN capability, PCM-CIA slots, a user-upgradable memory board, battery choices, a built-in data/fax modem, Ethernet compatibility, printer ports and removable 340-MB or 810-MB hard disk drives.

Gel analysis

Signal Analytics has announced an update to IPLab Gel, its **densitometry and gel analysis software** for the Macintosh (*Reader Service No. 105*). Version 1.5 contains easy-to-use commands specifically designed for performing densitometry and component sizing

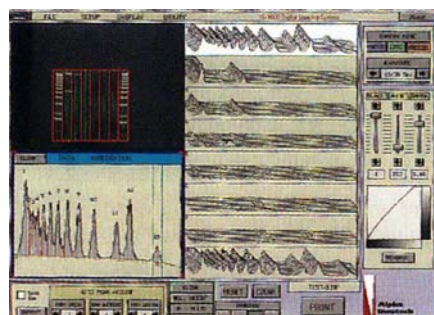


IPLab Gel for gel documentation.

on electrophoretic gel images. In addition to its previous capabilities for one- and two-dimensional density analysis on electrophoretic gel images, the program can determine molecular weights associated with bands within a gel image. Molecular weights are determined by matching band positions against user-defined standards. It now features special tools for drawing moveable objects into colour overlays. The objects can be used to define bands of interest for analysis,

and to annotate gel images for publication and documentation. A new export command saves annotated images as PICT files that preserve the high-resolution overlay drawings. IPLab Gel reads standard PICT and TIFF files, as well as Molecular Dynamics' PhosphorImager data files.

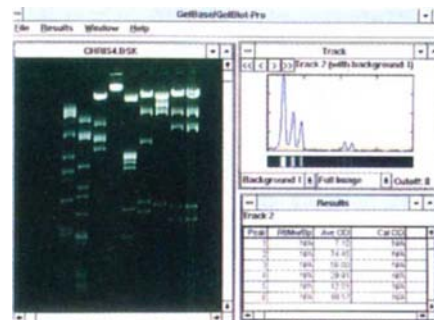
The IA-200 **image analysis software**, from Alpha Innotech, is designed to provide quick and comprehensive analytical capabilities for electrophoresis, films and membranes in a package offering a user-friendly, intuitive interface (*Reader Service No. 106*). Importing 8-bit TIFF, BMP, GIF



Alpha Innotech image analysis software.

or PCX file formats, the software performs densitometry, molecular weight calculations, RF measurements, dot/blot microtitre plate analysis, automatic colony counting, and quantitative PCR in less than three minutes. Results can be printed on a laserjet printer or exported as ASCII file formats compatible with spreadsheet programs.

Ultra Violet Products has introduced a new generation of **software for its gel documentation systems** (*Reader Service No. 107*). GelBase Pro Series software comprises a number of modular packages and is designed to analyse images from DNA/RNA and protein gels, autoradiographs, spots and blots. Routine analysis is said to be accurate and can be performed in less than three minutes — all software programs operate through a series of standardized windows and menu options. This package operates under both the PC and Apple Mac-based platforms. Users can start with GelBase Pro, a



UIVP's gel documentation software.

one-dimensional track analysis package, and add new software modules as their requirements grow. Programs for two-dimensional spot and blot analysis (GelBlot Pro), gel sequencing (GelSeq) and matching (GelMatch) are available. Images captured by the ImageStore 5000 or GDS7500 are transferred to the computer by floppy disk or by computer network, if a networking version of the ImageStore is in use. Alternatively, to eliminate disk transfer, the computer can be linked directly to the documentation system and a compatible frame-grabbing card ('Grab-It') fitted to the computer.

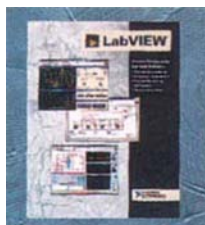
Data acquisition

Anville Instruments states that it has improved its successful series 410 range of low-cost, high-accuracy, **data-logging systems** by providing greater hardware and software facilities (*Reader Service No. 108*). The only extra equipment needed for any application are the sensors and a PC computer with Windows 3.1. It is said to be easy to configure — select the sensor type for each input individually (dc V; dc mA; PT100; types T, J, N, K, R, S, B and E thermocouples; pressure transducer) and the series 410 will configure itself with the correct signal conditioning. Each series 410 has eight analog inputs, eight digital inputs, one pulse-count input and eight digital outputs as standard — an option of two analog outputs is also available. The LOG410 software provides system configuration with individual input identification, scaling, calibration and alarms. The real-time graphical displays can be manipulated by the operator and all data are stored in compact date/time stamped files. A dynamic data link (DDE) to Microsoft Excel is standard.

Cambridge Electronic Design is launching the micro1401 **data-acquisition interface for recording experimental data** with a PC or Macintosh host computer (*Reader Service No. 109*). The micro1401 is approximately the size of a notebook PC and features its own RISC CPU and on-board memory. The design enables the unit to capture and analyse both waveform data and digital events with simultaneous analog- and digital-output sequencing capability. Waveform capture is available on four channels at 12-bit resolution. High-speed sampling is facilitated by a 333 kHz analog-to-digital conversion. Supplementary modules will be available to enhance the system's capabilities, including an additional 12-channel waveform input module and a digital in/out expansion module. The company has produced a range of software packages that enable the system to be customized for many life science applications, including intracellular and extracellular neurophysiology, EEG, evoked response and cardiovascular studies.

Imaging

National Instruments has announced the availability of a new, 16-page, full-colour LabView brochure (*Reader Service No. 110*). Entitled 'Graphical Programming for Instrumentation', it contains information on how scientists and engineers can use the program in test, measurement, process monitoring and control applications. The brochure includes examples of real-world applications developed with the software. It explains the front panel/block diagram hierarchy of the product, as well as the various virtual instrument libraries, add-on tool kits and other user-productivity tools. The brochure also lists additional resources available to users, such as user groups, an Internet forum for LabView users, technical support, the Alliance Program, and the Instrument Library Developer Program. The program is compatible with Windows, Windows/NT or Windows 95 PCs; Macintosh and Power Macintosh computers; Sun SPARCstations; and Hewlett-Packard 9000 Series 700 workstations.



LabView brochure.

Capturing and processing images is said to be easy with Data Translations' Vision-EZ (*Reader Service No. 111*). These **frame-grabber boards and software** for capturing, saving and printing images are designed to be an affordable and precise solution for imaging. Further expanding its product range, the company has introduced its lowest cost frame grabber, the Vision-EZ-LC. The IBM PC AT-compatible board is a monochrome frame grabber intended for cost-conscious applications. Included is the necessary start-up software, Global Lab Acquire, which is a subset of the company's established Image applications. The board captures images from any CCIR-compatible video source in real time with a resolution of 768 × 512 pixels and 256 grey levels. An external trigger input permits synchronization of acquisition with an external event. Vision-EZ hardware also ships with a simple DOS application and source. The company also offers Image and Image Development Environment software, available at extra cost for customers who wish to use or integrate a suite of image-analysis tools, including measurement, filtering and frequency analysis.

Optimization

Westing Software has announced the release of Card, the **computer-aided experiment design package** for Windows (*Reader Service No. 112*). The package was created for scientists and engineers working in product, process and quality

research and development. It is designed to move the user from risky trial-and-error to information-based experimenting. The system is designed to save time and money by reducing the number of experiments needed to find the knowledge and answers that yield optimized products and processes. Card includes a navigator that guides a novice user to the right experimental design for each situation. Built-in error trapping is said virtually to eliminate the possibility of operating mistakes. The program can be used to explore and quickly recover from flaws or gaps in historical or plant data. These data are easily input and then analysed for any information problems, the program then



Card computer-aided design package.

creates a recovery design that corrects the problems with a minimum of work required. Analysis for Answers, a companion product, is designed to make analysis of experiments quick and easy — Card automatically creates analysis-ready data matrices. Correctly structured and formatted for numerical analysis, they can also be exported in a variety of file formats compatible with virtually all spreadsheet and statistics packages.

The latest development in Perkin-Elmer's SQL LIMS is an innovative user interface. SQL LIMS server software is based on Oracle Forms 4.5 and is implemented on the Windows interface (*Reader Service No. 113*). The SQL LIMS server software is designed to maximize the power and open standards of the Oracle V7 relational database and CDE 2 tools. It is available for use with Hewlett-Packard's HP-UX, Digital's VAX and AXP, and platforms from SUN Microsystems. SQL LIMS is said to provide superior information management tools and optimum information processing facilities for the analytical laboratory. It automates data acquisition, processing and archiving functions, and provides management information to assist in the efficient planning and control of laboratory resources.

Net news and CD-ROMs

Environmental Systems Research Institute (ESRI) offers a comprehensive **home page**, which can be accessed on the World Wide Web users using NCSA Mosaic, Netscape, Spyglass, or other browser soft-

ware: <http://www.esri.com> (*Reader Service No. 114*). ESRI's home page consists of hundreds of unique documents organized into ten general areas. Among many other items, network 'surfers' will find a worldwide company directory, GIS news and articles, software product information and numerous technical resources. Several free items are offered, including ESRI's special offer of a no-cost licensed version of ArcView Version 1.0 for Windows software available for downloading. Several geographic data sets are also available to browsers, many for use with ArcView Version 1.0 for Windows. In addition, a variety of maps and images can be viewed, printed or downloaded.

Microinfo now offers **CD-ROM products and other information media** for professional applications — some 40 databases on subjects as varied as aquatic sciences, biomedicine, biotechnology, the environment and pollution, genetics, neurobiology and zoology are available (*Reader Service No. 115*). All of these databases are available either on magnetic tape for downloading and searching on customer-owned computer systems or directly from a host served by means of Internet access. Payment of a single subscription for each database of interest allows users unlimited access to that database for a period 12-month. All databases are updated monthly for both magnetic tape and Internet options. Databases in this series are prepared and published by Cambridge Scientific Abstracts (CSA) with the exception of certain titles, such as UN Specialized Agencies, the US National Library of Medicine, Engineering Information or Materials Information (ASM International/Institute of Metals). Prospective subscribers are invited to use any CSA Internet database service free of charge for up to 30 days. □

These notes are compiled by Brendan Horton from information provided by the manufacturers. For more details, fill in the reader service card bound inside the journal.

ADVERTISEMENT

OLIGONUCLEOTIDES

Synthesised, purified and shipped in 3 days

£2.50 per base

(All inclusive)

0.2 µM scale (approx. 500 µg)

- Antisense (completely or partially sulphurised)
- Extensive range of labelling and modifications
- Full design and analysis facilities
- Gene cassettes, bulk order and contract discounts

Molecular Medicine Unit
 King's College School of Medicine & Dentistry
 Tel: 071 346 3126 Fax: 071-733 3877

Reader Service No.10